

AI4S对科学研究程序的影响

——以AlphaFold2为例

The Impact of AI for Science on Scientific Research Procedures: A Case Study of AlphaFold2

祁丽萍 / QI Liping 张增一 / ZHANG Zengyi

(中国科学院大学人文学院, 北京, 100049)
(School of Humanities, University of Chinese Academy of Sciences, Beijing, 100049)

摘要: 文章从分析AlphaFold2的科学研究过程入手, 对其模型架构和研究程序进行了深度剖析。AlphaFold2首先构建蛋白质序列和结构的数据表征, 在此基础上进行模型训练, 并基于蛋白质结构的相关知识进行机器猜想, 然后利用迭代循环验证、修正蛋白质三维结构并输出结果。AlphaFold2给科学研究程序带来的影响主要体现在以下三个方面: 首先形成了跨领域的人机协同科学研究主体网络; 其次, 科学研究对象从蛋白质实体转向氨基酸数据; 第三, 研究手段从实验室实验转向深度神经网络模型开发。然而AlphaFold2本身也存在诸多局限包括不透明性和不可解释性问题。研究旨在为探讨AI for Science对科学研究程序带来的影响提供一个具体案例。

关键词: 人工智能 AlphaFold2 科学研究 科学方法

Abstract: This paper begins with an analysis of AlphaFold2's research process, offering an in-depth examination of its model architecture and research procedures. AlphaFold2 first constructs data representations of protein sequences and structures, upon which model training is conducted; it then generates machine-derived hypotheses grounded in knowledge of protein structures, which is followed by an iterative revision of three-dimensional protein structures and result output. The impact of AlphaFold2 on scientific research procedures is primarily in three aspects: first, the formation of a cross-disciplinary, human-machine collaborative research network; second, a shift in the object of scientific inquiry from physical protein entities to amino acid data; and third, a transformation in research methodology from laboratory experimentation to the development of deep neural network models. Nevertheless, AlphaFold2 is subject to several inherent limitations, including opacity and inexplicability. This paper provides a concrete case study for examining the broader implications of AI for Science on scientific research procedures.

Key Words: Artificial intelligence; AlphaFold2; Scientific research; Scientific method

中图分类号: TP18; N031 DOI: 10.15994/j.1000-0763.2026.08.003 CSTR: 32281.14.jdn.2026.08.003

基金项目: 中国科协2024年度研究生科普能力提升项目“人工智能驱动科普智库研究范式变革研究”(项目编号: KXYJS2024003)。

收稿日期: 2025年2月10日

作者简介: 祁丽萍(1994-)女, 湖北云梦人, 中国科学院大学人文学院博士研究生, 研究方向为人工智能与社会、科学传播。Email: qiliping22@mailsucas.ac.cn

张增一(1964-)男, 山东聊城人, 中国科学院大学人文学院教授, 研究方向为科学传播、科技方法论、科技与社会。Email: zhzy@ucas.ac.cn

引言

随着大数据、新算法以及人工智能技术的创新发展并在科学领域的广泛应用,人工智能驱动科学研究(AI for Science, AI4S)引起了社会各界的关注。目前,中国、美国、英国、法国等国都在积极部署推动人工智能(以下简称“AI”)与科学的结合,^[1]如使用AI技术来加速新材料的设计、发现和评估,推动自动化实验室的开发,帮助实现药学模型的建立和优化、药物筛选或药效预测等。^{[2], [3]}越来越多的研究者认识到AI在科研中的潜力^[4]及其对科学研究模式的影响。^{[5], [6]}斯蒂凡诺·毕安齐尼(Stefano Bianchini)认为AI将“重塑发现过程的本质并影响科学的组织结构”。^[7]

AlphaFold2作为最具代表性的AI for Science案例,被美国《科学》杂志列为“2021年度十大科学突破”榜首,并获得2024年诺贝尔化学奖。^[8]AlphaFold2利用基于注意力机制的深度学习算法,对大量蛋白质序列和结构数据进行训练,并结合物理学、化学和生物学的先验知识,直接训练蛋白质结构的原子坐标,实现端到端的蛋白质结构预测。^[9]因此,本文以AlphaFold2为案例,尝试回答以下问题:AI驱动的科学在程序和方法上与传统科学研究有何不同?AI对科学研究程序带来了哪些新变化?

一、蛋白质结构预测方法的演变

蛋白质的三维结构在很大程度上决定了其功能,了解这些结构和功能对于药物开发、疾病治疗以及洞察生命如何运作等至关重要。^[10]蛋白质结构测定的方法经历了从实验解析方法到计算预测方法的变迁,深刻体现了科学研究中从规则构建到数据驱动的逻辑转向。

1. 实验方法

为了测定蛋白质的结构,生物学家们早期采用的是直接观测和实验方法,包括X射线晶体衍射(X-ray Crystallography)、核磁

共振光谱法(Nuclear Magnetic Resonance, NMR)和冷冻电子显微镜方法(Cryo-Electron Microscopy, Cryo-EM)。

X射线晶体衍射作为最早解析蛋白质结构的方法,首先将蛋白质进行结晶化处理,然后将蛋白质晶体固定在X射线束路径中使用高能X射线照射晶体并产生衍射图样,使用相位和强度信息构建蛋白质晶体的电子密度图,最后根据电子密度图构建蛋白质的初步原子模型重建蛋白质的三维结构。^[11]1958年,约翰·肯德鲁(John Kendrew)和马克斯·佩鲁茨(Max Perutz)用X射线晶体衍射法首次解析出了血红蛋白和肌红蛋白的三维结构。^[12]该方法虽然准确解析了蛋白质结构,但是具有较大的局限性。一是依赖高质量的蛋白质晶体,但是许多蛋白质难以结晶;二是十分耗时,首次解析出血红蛋白和肌红蛋白的三维结构花费了20多年的时间。

为了解析自然状态下的蛋白质结构,1975年库尔特·伍特里希(Kurt Wüthrich)开发了利用核磁共振技术来测定溶液中生物大分子三维结构的方法,解析了液态蛋白质结构,并获得了2002年的诺贝尔化学奖。^[13]核磁共振光谱测定蛋白质结构的原理是基于原子核在磁场中的自旋特性,氢原子等带有磁矩的核在外加磁场下会产生共振,通过检测这种共振信号,可以推测原子的空间位置和相互作用。该方法解决了蛋白质结晶难题,可解析蛋白质在溶液中的动态信息,得到的结构更接近蛋白质的天然状态。但是该方法对蛋白质的纯度有极其高的要求,数据处理更为复杂,测定时间仍然较长,并且仅适用于解析小型蛋白结构,大型蛋白解析难度大。

为了解析大分子复合体以及大型蛋白质结构,随后出现了冷冻电子显微镜技术,雅克·迪波什(Jacques Dubochet)等人首次提出冷冻样品技术,使得生物分子样品在冷冻状态下免受辐射损伤,保持其天然结构。^[14]冷冻电镜通过在极低温条件下(通常低于-150℃)快速冷冻生物样品,保持其接近天然的结构状态,然后利用电子显微镜获取样品的大量二维投影图

像, 并通过计算机算法将这些图像重建为三维结构, 但是该方法一直受到分辨率不足的限制。^[15]随着电镜设备和数据处理算法的进步, 冷冻电镜的分辨率显著提高, 2017年获诺贝尔化学奖认可的冷冻电镜技术成为解析大分子复合体和细胞器结构的重要手段。该方法的实验条件要求苛刻, 设备也十分昂贵且仍然具有较大的计算成本。

综上所述, 在长达数十年的时间里, 蛋白质三维结构主要依赖高精度实验手段获得。然而这些实验方法普遍存在操作复杂、设备昂贵、周期漫长、适用范围有限等问题。尤其是在基因组测序技术推动下, 蛋白质序列数量呈指数增长, 而实验解析结构的能力远远无法匹配, 几十年来利用传统的实验方法确定了大约10万个独特蛋白质的结构, 但这仅代表数十亿已知蛋白质序列的一小部分, 从而形成了“序列-结构”鸿沟。^[16]20世纪70年代, 克里斯蒂安-安芬森(Christian Anfinsen)正式提出著名的“安芬森法则”即蛋白质的一维序列决定蛋白质的三维结构, 为蛋白质的结构预测方法提供了理论支撑。^[17]正是在这种背景下, 根据氨基酸序列确定蛋白质的三维结构的预测问题成为关键研究方向, 由此开启了计算建模主导的结构预测方法。^[18]

2. 计算方法

结构生物学家从20世纪70年代就开始尝试计算方法预测蛋白质结构。最早的计算方法主要基于物理和化学规则, 包括同源建模、折叠识别^[19]和从头预测(Abinitio Prediction)^[20]三种路径。这些方法的基本理论是基于蛋白质结构的保守性以及蛋白质的共进化规律, 可通过序列比对找到相似模板, 再通过能量函数优化构建三维模型。尽管这些方法在一定程度上推动了结构预测进展, 但其核心问题在于高度依赖已有模板数据库、能量函数构建困难、搜索空间巨大等。尤其在缺乏已知结构模板的情况下, 这些方法难以提供可靠预测。这一时期的计算方法整体上体现出一种规则驱动的思路, 试图用人类可理解的物理模型或知识规则, 去还原蛋白质折叠机制。

AI的发展, 特别是神经网络在图像识别与自然语言处理领域取得的突破, 为蛋白质结构预测带来全新的解决思路。深度学习方法不再依赖人工构建的能量函数或显式规则, 而是依靠大规模数据学习蛋白质序列与结构之间的复杂映射关系。AlphaFold2不再试图还原蛋白质折叠的物理机制, 而是专注于以最大限度的精度逼近结构结果本身。它不再追求解释蛋白质折叠的自然过程, 而是借助数据和模型从相关性出发实现可预测性, 标志着结构生物学从物理建模转向神经网络建模。深度学习的引入彻底改变了蛋白质结构预测的格局, 尤其是在AlphaFold、RoseTTAFold等模型的推动下, 深度学习已成为蛋白质结构预测的主流方法。在CASP14中, DeepMind公司开发的AlphaFold2在全局距离测试中突破了90分, 其利用深度学习方法直接进行端到端的训练, 对几乎所有的蛋白质都预测出了正确的拓扑结构。AlphaFold2显著扩大了蛋白质组的结构覆盖范围, 其规模几乎覆盖了整个人类蛋白质组(98.5%的人类蛋白质)。^[21]

综上所述, 蛋白质结构预测方法的演变经历了实验精度优先→计算效率优先→预测精度优先的过程。在传统实验手段难以应对规模性需求的背景下产生了计算方法, 然而计算方法初期受限于建模能力与规则复杂性。深度学习方法运用各种算法规则, 对模型迭代优化不断探索蛋白质结构的潜在规律, 最终实现了原子级精度的AI方法预测。AlphaFold2的出现, 使蛋白质结构预测不再只是结构生物学内部的问题, 而成为AI、数据科学与生命科学深度交叉的新平台, 推动科学研究程序的全面变革。

二、AlphaFold2的科学研究过程

AlphaFold2本质上是一个基于Transformer架构的具有庞大参数的深度神经网络模型, 从组织架构来看, AlphaFold2可以分为三大模块, 输入模块、主干模块和循环模块, 利用数据+深度神经网络模型+新型算法+科学知识, 实现AI在科学研究中的应用。

1. 构建蛋白质序列和结构的数据表征

AlphaFold2 中的两个关键数据表征是多序列比对 (Multi-Sequence Alignments, MSA) 和对表征 (Pair Representation)。输入模块以蛋白质的氨基酸序列作为信息输入, 将氨基酸序列在遗传序列数据库中进行搜索并构建多序列比对, 可以得出蛋白质序列的共进化信息和保守性信息, 在此基础上计算序列的关键特征。同时 AlphaFold2 从蛋白质结构数据库中搜索并构建表示残基之间关系的对表征。对表征用于表示每两个蛋白质残基之间的相关性, 通过相关性信息提取两个氨基酸之间的距离特征。通过这种方式, 一维氨基酸序列通过领域知识和数据集进行扩充, 形成二维关系表征。

如萨比娜·莱奥内利 (Sabina Leonelli) 认为数据是经验知识的合法基础, 数据的产生相当于捕获世界的特征, 可以用于系统研究。^[22] AlphaFold2 拥有足够多的序列数据从而得到良好的多序列比对结果, 可以从序列中直接学习氨基酸之间的相互作用关系。通过这两个数据表征学习大量蛋白质的序列和结构数据, 从中寻找氨基酸分子之间的相互作用, 以及蛋白质片段之间的演化关系, 然后再按照找到的规律对蛋白质的结构进行预测。

2. 利用蛋白质序列数据进行模型训练

AlphaFold2 同时使用带标签和未带标签的数据进行模型训练, 并通过自蒸馏的方式使用未标注数据来弥补标注数据的不足。已标记的数据就是蛋白质结构数据库 (Protein Structure Database, PDB), 利用原始的 PDB 中的 15 万结构数据进行训练, 得到初步的预测模型。然后使用这个模型对未标记的数据库 UniClust30 的约 35 万个不同序列的结构进行预测, 并制作一个新的预测结构数据集对高质量的预测进行筛选, 再将其放回训练集中构建高置信度训练子集。随后使用 PDB 数据和这个新的预测结构数据集作为训练数据的混合物, 再次从头开始训练相同的架构。^[9]

这种自蒸馏过程有效地利用了未标记的序列数据, 并显著提高了所得结果的准确性。使用 PDB 和新的预测蛋白质结构的自蒸

馏未标记数据集的组合作为训练数据用来训练 AlphaFold2, 目的是通过使用不同的训练数据增强方法, 大大提高了模型的性能, 实现对大量氨基酸序列、蛋白质结构等观察数据的准确拟合。

3. 基于蛋白质结构知识的机器猜想

研究团队将蛋白质知识放入模型中, 利用蛋白质结构的进化、物理和几何约束原理来指导蛋白质结构的预测, 以提高预测的准确性和生物学合理性。例如蛋白质结构的保守性是折叠识别的理论依据, AlphaFold2 通过分析蛋白质序列的进化信息, 如保守性分析和同源序列比对, 来推断蛋白质结构的可能结构域和相对位置。^[23] 通过分析同源序列中的共变, 可以推断出哪些氨基酸残基处于接触状态, 这有助于蛋白质结构的预测。

Evoformer 作为结构预测的注意力架构, 充分混合更新对表征和 MSA 之间的信息, 结构模块将更新后的 MSA 和对表征信息嵌入三维空间, 利用三维空间结构平移和旋转等变的知识, 引入不动点注意力 (Invariant Point Attention, IPA) 计算蛋白质结构的全原子坐标, 结构模块同时对主链结构和侧支链结构进行优化, 实现原子水平的端到端的三维结构预测和优化。

4. 利用迭代循环验证、修正蛋白质三维结构并输出结果

AlphaFold2 团队通过引入迭代算法, 采用了多轮循环迭代优化, 根据训练数据的结果不断对模型进行微调, 从而找出最合适的模型。多轮迭代让模型深度更深, 通过迭代让模型预测更精确, 可以预测到更复杂的结构。AlphaFold2 利用机器猜想完成了智能机器试错, 根据机器学习 AlphaFold2 初步构想蛋白质的三维结构, 然后利用算法和 Transformer 架构等实现关键表征信息的更新和交互, 加上循环迭代机制, 对初步生成的蛋白质三维结构进行修正和调整。最后 AlphaFold2 的多输出, 对模型本身预测的准确度进行打分。为了使预测结果具有实际意义, 对预测结果进行良好的校准和序列解析置信度测量。AlphaFold 团队构建了置信度指标对 AlphaFold2 的输出结果进行检验。

综上所述, AlphaFold2的研究过程可分为三个主要部分。第一步, 系统将与输入序列相对应的MSA信息、序列信息和模板信息嵌入MSA表征和对表征。第二步, 利用由两个变换器组成的Evoformer更新这两个表征信息。第三步结构模块根据Evoformer最后更新的输出表示估算三维结构。此外, 在AlphaFold2中, 循环模块应用于整个神经网络中, 通过反复将最终损失应用于输出, 然后将输出递归地反馈到相同的模块中, 进而实现三次强化迭代优化。

三、AlphaFold2对科学研究程序的影响

李正风指出要理解当代科学的新变化可以从两个方面入手, 一是研究对象的新变化, 二是研究方法的新变化。^[24] AlphaFold2作为人工智能神经网络模型初步解决了蛋白质结构预测问题。在传统的蛋白质三维结构研究中, 科学家需要哪个蛋白质的三维结构, 就开始对该蛋白质进行观察、实验测定。这个研究过程是对自然对象的观察, 然后根据蛋白质的性质, 采用合适的实验方法。AlphaFold2利用深度学习方法, 解决的是如何用计算方法解决蛋白质三维结构预测的问题。那么AI对科学研究程序的影响主要体现在以下几个方面。第一, 科学家的跨领域合作和人机协同成为重要的科学研究主体。第二, 研究对象从蛋白质实体转向氨基酸数据。第三, 研究手段从实验室实验转向深度神经网络模型开发。此外, AlphaFold2本身仍存在一定的局限性。

1. 人-人与人-机成为科学研究的新主体

科学研究的主体往往被认为局限在相应的科学领域, 然而AI时代打破了学科之间的界限, 使得AI专家、企业和非学术机构都可以参与到基础科学研究中, 实现了人-人的跨领域合作。

AlphaFold2成功实现了科学家与AI专家的跨领域合作。从AlphaFold2的研究团队来看, 共有34名研究人员, DeepMind公司的研究员有33位, 其中有一名研究员来自于韩国首尔国立大学生物科学学院和人工智能研究所。在作者贡献度排名中, 他们列出了19位研究团队

的核心成员, 表明他们对于AlphaFold2的开发做出了同等重要的贡献。他们的团队成员拥有高度跨学科的性质, 有生物学家、化学家和物理学家, 还有机器学习和工程学专家, 成功实现了科学、计算机、工程学的跨领域合作。^[9] DeepMind的创始人戴米斯·哈萨比斯(Demis Hassabis)也是AlphaFold2项目的构想者, 具有神经科学和AI的双重背景。整个AlphaFold2项目的完成主要分为项目构思、开发神经网络架构和训练、开发数据、分析和推理系统、创建最大的公开蛋白质家族数据库(Big Fantastic Database, BFD)、基因组学数据库、构想端到端方法、工程技术协助等等板块, 集合了科学家、AI专家、工程师等研究人员。

AI作为智能体与人类共同参与科学研究, 从AI4S全过程的三个阶段来看, 大模型建构、大模型运行和对大模型生成结果的解释, 这三个阶段都实现了人机协同。一方面, 大模型的建构和运行是人与AI系统不断互动的过程。研究者根据科学目标将相关领域的专业知识嵌入深度学习算法中, 即让AI系统对先验知识进行学习, 实现物理世界与数学模型的拟合。在AlphaFold2项目中, 要将蛋白质知识嵌入AI神经网络模型中, 包括蛋白质结构的进化、物理和几何约束等先验知识, 经过这些大量的数据训练后建构针对特定问题蛋白质三维结构预测的经验模型, 保证模型预测的准确率以及预测结果符合相关领域的运行规律, 不违反常识。大模型的运行阶段也是深度学习的一部分, 由于深度神经网络的结构通常是不透明的, 因此研究者需要结合科学目标以及AI生成的内容不断进行神经网络架构和算法的调整, 直到生成符合预期的内容。所以大模型的建构和运行是密不可分相辅相成的两个阶段, 而且需要研究者在其中发挥主导作用, 对大模型的准确性进行把握。

另一方面, 对神经网络模型生成结果的解释需要由人来进行。AlphaFold2预测的蛋白质结构之所以能够实现原子级精度并且具有可信度正是因为研究者对AlphaFold2生成结果的解释和验证。AlphaFold2团队建构了置信度指标

来衡量每个残基位置的置信度,神经网络模型在生成结果的同时,还会生成衡量结果可信度的参数,这个生成结果以及置信度指标由研究者进行解读。如果没有置信度指标,仅凭研究者的直觉或者经验也很难对结果进行判断和解释。因此,对该模型输出结果的解释也是人机协同的过程。

将科学家的经验知识和创造力与深度神经网络模型相结合,科学家可以通过AI快速获取并分析海量数据,并且建构用于科学研究的深度神经网络模型,从而发现新的规律和趋势进行科学预测,这也意味着科学家与AI在科学研究实践的人机协同探索取得进展。

2. 研究对象从蛋白质实体转向氨基酸数据

传统蛋白质结构测定以物质形态的蛋白质本身作为直接研究对象。X射线晶体学、冷冻电镜与核磁共振是三种主流实验手段,各自通过不同的物理原理对蛋白质样品施加处理。X射线晶体学长期以来是生物大分子结构测定的主导技术,其基本流程是将蛋白质样品制备成晶体,以高能X射线照射后收集衍射图样,再从中提取振幅信息,结合相位估算重建三维结构。冷冻电镜则无需结晶,将样品快速冷冻后以电子束成像。上述方法尽管技术原理各异,但共同点在于研究者必须取得实物蛋白质样品并对其实施物理处理,最终结构信息从实验信号中提取而来。

AlphaFold2的出现使研究对象发生了转变,直接处理的不再是物质形态的蛋白质,而是氨基酸序列数据与已知蛋白质结构数据。AlphaFold2以氨基酸序列为输入,在多个蛋白质序列数据库中构建多序列比对,以判断序列中哪些位置容易发生突变,并检测位点之间的协同进化关联;同时检索是否存在具有已知三维结构的同源蛋白,以此构建氨基酸对的初始表征。在训练数据方面,AlphaFold2所使用的序列数据库包括UniRef90、UniClust30等大规模蛋白质序列库,并且利用海量的宏基因组数据进行隐马尔可夫模型建模构建了BFD。^[9]模型从氨基酸序列出发,通过MSA提取序列的协同进化信息,并构建氨基酸间的成对距

离矩阵,进而预测蛋白质三维结构。换言之,AlphaFold2所寻找的是序列数据与结构数据之间的统计对应规律,而非通过物理实验直接探测蛋白质分子。

两种路径的核心差异在于研究对象的性质及知识的获取方式。传统实验方法以物质实体为对象,每一个结构的测定对应一次独立的实验操作,结构信息直接来自对蛋白质样品的物理测量,因而具有明确的因果链条,样品制备、信号采集、模型重建,每一步骤均有对应的物理依据。然而这一过程代价高昂,测定单个蛋白质结构往往需要数月乃至数年的工作,导致已知序列中仅有极小一部分获得了实验确定的结构。AlphaFold2则以序列与结构的数据对应关系为操作对象,其预测能力来自对大量历史实验数据中隐含规律的提取,而非针对具体蛋白质实施新的实验。AlphaFold2的预测精度达到原子级别,这一性能跃升远超学界预期。这意味着,AlphaFold2的研究对象在认识论意义上是数据层面的规律性,而非物质层面的具体分子,知识的获取途径从直接的物理测量转向了对既有数据的模式提取。

3. 研究手段从实验室实验转向深度神经网络模型开发

蛋白质三维结构的传统测定依赖于在实验室中对蛋白质样品直接施加物理操作。X射线晶体学是一套从基因到结构的完整实验流程,涵盖蛋白质结晶、数据采集、数据处理、结构解析与精修等多个环节。冷冻电镜与NMR虽然无需结晶,但同样要求制备高纯度的蛋白质样品并实施复杂的仪器操作。研究者必须同时掌握实验观测数据与关于分子结构的已有知识,方能构建最终的原子模型。这一过程耗时漫长,单个蛋白质结构的解析往往历时数月乃至数年。

AlphaFold2作为一种深度神经网络模型,实现了实验水平的蛋白质结构预测,在科学实践领域产生了重要影响,使得建构深度神经网络模型成为一种全新的研究手段,是利用深度学习等算法建构科学领域的神经网络模型,对大量的科学知识和实验数据进行学习,能够依

据科研目标直接生成需要的内容。这种科学研究从传统的实验室研究转向为一项“复杂的社会工程”，AI科学大模型的知识生产方式也将科学问题与数学问题、计算机建模问题以及工程问题等结合起来。^[25]

AlphaFold2所采用的研究手段，是开发和训练深度神经网络模型，实现了“大数据”“算法”“知识”和“模型”的完美融合。在大数据方面，PDB中实验蛋白质结构的稳定增长、基因组测序的爆炸式增长为多序列比对提供了丰富的实验数据。此外，要让AlphaFold2寻找蛋白质结构数据背后蛋白质结构的规律，需要将蛋白质结构的物理和生物学知识结合起来，这个过程整合了生物学家的发现，以及来自遗传学、数学和化学领域的知识。再者，通过大数据与多种算法、科学知识的融合建构了解决蛋白质结构问题的神经网络模型，共同驱动AI参与科学研究。因此，AlphaFold2的研究过程的核心工作是神经网络架构的设计、训练策略的制定与参数的优化。

目前在化学的每个子领域都开发了各种形式的深度神经网络模型，应用于化学合成、催化和材料发现等方面。^[26]2023年11月谷歌DeepMind团队发布了材料探索图形网络模型（Graph Networks for Materials Exploration, GNoME），通过大规模主动学习预测并验证了220万种潜在晶体结构，新增38.1万种至材料项目数据库，发现42.1万个稳定晶体，显著扩展了材料发现的边界。^[27]阿尔特·泰兰（Altae-Tran）团队基于深度聚类算法FLSHclust识别出200余种新型CRISPR相关功能系统，涵盖13万个相关基因，其中188个此前未知，为微生物蛋白质功能多样性研究提供了新路径。^[28]

由此可见，利用AI构建深度神经网络科学模型使得科学转向规模化的知识生产模式，大大提高了科学研究效率，为解决目前科学界的很多难题提供了一种有效的方法和手段。^[29]科学家和科研工作者能够从重复性的劳动中解放出来，投入更多的时间和精力放在更有创造力的工作中，提出更多有价值的问题，解决人类面临的更紧迫的科学问题。这对于推动科学的

发展，科学进步具有重要作用。

4. AlphaFold2的局限性

AlphaFold2的突破性成就表明AI在未来科学研究中拥有巨大潜力，^[30]然而也存在诸多局限包括对科学研究过程的透明性与研究结果的可解释性问题。

联合国2021年发布的首个全球性人工智能伦理协议《人工智能伦理问题建议书》，提出AI系统应当遵循的十个原则包括“透明度和可解释性”。^[31]透明性原则要求科学研究的过程公开透明，强调实验数据和结果的可重复性，^[32]同行可以重复科学研究的完整过程。可解释性寻求对科学结果的因果解释和清晰的推理过程，传统科学“将寻求解释作为科学的基本目标”。^[33]科学界原就存在可重复性危机，发表在学术期刊上的论文很多都不能依据研究过程重现，^[34]AI的引入进一步挑战了科学的透明性。

AlphaFold2的训练过程是黑箱状态，黑箱效应指深度神经网络、大语言模型等复杂算法系统依赖海量参数、非线性结构与多层训练过程，这使得输入与输出之间的因果路径无法被清晰揭示，模型内部的决策逻辑也难以被研究者所理解。^[35]例如，AlphaFold2能够实现原子级精度的蛋白质三维结构预测，但是它却不能告诉科学家蛋白质的折叠过程和解释折叠规律。透明性和可解释性问题密切相关，解决可解释性危机的前提就是增强透明性，这两者都直接影响着AI的可信度，不可信的科学结果会破坏未来研究的基础，从而阻碍科学进步。^[36]

总 结

综上所述，AlphaFold2作为AI4S的经典案例，形成了一种新的科学研究程序路径。原始科学数据经由算法处理转化为模型的训练数据，模型在领域知识的约束下通过参数优化生成科学预测，预测结果反馈至数据收集与模型修正的新一轮循环，科学知识在这一持续迭代的闭环中不断深化。科学研究程序不再是由单一线性逻辑贯穿的顺序流程，而是由多要素协同驱动的动态迭代系统，程序的推进动力从人

类的理论直觉转向数据与模型的交互优化。

虽然 AlphaFold2 具有一定的局限性,但是有学者指出如果一个模型有产生正确输出的历史,我们就有理由相信它。^[37]也有学者认为 AI 模型并不需要可解释性才有用。^[38]就像在医学领域,医生通常在不知道治疗方法如何或为什么有效的情况下提供治疗方案,也就是说只要大模型能够产生好的结果我们就有理由使用 and 相信它。因此在 AI 的参与下,我们应该聚焦于科学研究进入新的发展阶段,产生了哪些新的研究方法和手段,拓展和丰富科学方法论的内涵。

【参考文献】

- [1] 杨小康、许岩岩、陈露. AI for Science: 智能化科学设施变革基础研究 [J]. 中国科学院院刊, 2024, 39 (1): 59-69.
- [2] 凌曦、赵志刚、李新刚. 人工智能技术在药学领域的应用——基于 Web of Science 的文献可视化分析 [J]. 中国药房, 2019, 30 (4): 433-438.
- [3] Szymanski, N. J., Rendy, B., Fei, Y., et al. 'An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials' [J]. *Nature*, 2023, 624(7990): 86-91.
- [4] 杨金龙、江俊. 拥抱科研新范式——人工智能带来的科研革命 [J]. 科学与社会, 2023, 13 (3): 11-22.
- [5] 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.
- [6] 王飞跃、缪青海. 人工智能驱动的科学新范式: 从 AI4S 到智能科学 [J]. 中国科学院院刊, 2023, 38 (4): 536-540.
- [7] Bianchini, S., Müller, M., Pelletier, P. 'Artificial Intelligence in Science: An Emerging General Method of Invention' [J]. *Research Policy*, 2022, 51(10): 104604.
- [8] Service, R. 'Science's 2021 Breakthrough of the Year: AI Brings Protein Structures to All' [EB/OL]. *Science*, <https://www.science.org/content/article/breakthrough-2021>. 2021-12-16.
- [9] Jumper, J., Evans, R., Pritzel, A., et al. 'Highly Accurate Protein Structure Prediction with AlphaFold' [J]. *Nature*, 2021, 596(7873): 583-589.
- [10] Huang, Y., Jiang, J., Wu, M. 'Discovering Deaminases Using AlphaFold2: A Strategy to Search for Tool Proteins for Gene Editing' [J]. *Signal Transduction and Targeted Therapy*, 2024, 9(1): 1-3.
- [11] Kendrew, J. C. 'Myoglobin and the Structure of Proteins' [J]. *Science*, 1963, 139(3561): 1259-1266.
- [12] Kendrew, J. C., Bodo, G., Dintzis, H. M., et al. 'A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis' [J]. *Nature*, 1958, 181(4610): 662-666.
- [13] Wüthrich, K. 'NMR with Proteins and Nucleic Acids' [J]. *Europhysics News*, 1986, 17(1): 11-13.
- [14] Dubochet, J., Adrian, M., Chang, J. J., et al. 'Cryo-electron Microscopy of Vitri-fied Specimens' [J]. *Quarterly Reviews of Biophysics*, 1988, 21(2): 129-228.
- [15] Saplakoglu, Y. 'How AI Revolutionized Protein Science, but Didn't End It' [EB/OL]. *Quanta Magazine*, <https://www.quantamagazine.org/how-ai-revolutionized-protein-science-but-didnt-end-it-20240626/>. 2024-06-26.
- [16] PDB Consortium. 'Protein Data Bank: The Single Global Archive for 3D Macromolecular Structure Data' [J]. *Nucleic Acids Research*, 2019, 47(D1): D520-D528.
- [17] Anfinsen, C. B. 'Principles that Govern the Folding of Protein Chains' [J]. *Science*, 1973, 181(4096): 223-230.
- [18] Senior, A. W., Evans, R., Jumper, J., et al. 'Improved Protein Structure Prediction Using Potentials from Deep Learning' [J]. *Nature*, 2020, 577(7792): 706-710.
- [19] Sander, C., Schneider, R. 'Database of Homology-derived Protein Structures and the Structural Meaning of Sequence Alignment' [J]. *Proteins: Structure, Function, and Bioinformatics*, 1991, 9(1): 56-68.
- [20] Simons, K. T., Kooperberg, C., Huang, E., et al. 'Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences Using Simulated Annealing and Bayesian Scoring Functions' [J]. *Journal of Molecular Biology*, 1997, 268(1): 209-225.
- [21] Tunyasuvunakool, K., Adler, J., Wu, Z., et al. 'Highly Accurate Protein Structure Prediction for the Human Proteome' [J]. *Nature*, 2021, 596(7873): 590-596.
- [22] Leonelli, S. 'Scientific Research and Big Data' [EB/OL]. *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2025/entries/science-big-data/>. 2020-05-29.
- [23] Bouatta, N., Sorger, P., AlQuraishi, M. 'Protein Structure Prediction by AlphaFold2: Are Attention and Symmetries All You Need?' [J]. *Acta Crystallographica Section D: Structural Biology*, 2021, 77(8): 982-991.
- [24] 李正风. 当代科学的新变化与科学学的新趋向 [J]. 世界科学, 2024, (8): 41-44.

- [25] 贾向桐. 大数据的新经验主义进路及其问题[J]. 江西社会科学, 2017, 37 (12) : 5–11.
- [26] Xu, Y., Liu, X., Cao, X., et al. 'Artificial Intelligence: A Powerful Paradigm for Scientific Research'[J]. *The Innovation*, 2021, 2(4): 100179.
- [27] Merchant, A., Batzner, S., Schoenholz, S. S., et al. 'Scaling Deep Learning for Materials Discovery'[J]. *Nature*, 2023, 624(7990): 80–85.
- [28] Altae-Tran, H., Kannan, S., Suberski, A. J., et al. 'Uncovering the Functional Diversity of Rare CRISPR-Cas Systems with Deep Terascale Clustering'[J]. *Science*, 2023, 382(6673): eadi1910.
- [29] 王飞跃、缪青海、张军平. 探讨 AI for Science 的影响与意义: 现状与展望[J]. 智能科学与技术学报, 2023, 5 (1) : 1–6.
- [30] Birhane, A., Kasirzadeh, A., Leslie, D., et al. 'Science in the Age of Large Language Models'[J]. *Nature Reviews Physics*, 2023, 5(5): 277–280.
- [31] 联合国教科文组织. 人工智能伦理问题建议书[R]. 巴黎: 联合国教科文组织, 2022, 5.
- [32] Al-Khalili, J. 'The “First True Scientist”'[EB/OL]. BBC NEWS, <http://news.bbc.co.uk/2/hi/science/nature/7810846.stm>. 2009–01–04.
- [33] Maxwell, N. 'A Critique of Popper's Views on Scientific Method'[J]. *Philosophy of Science*, 1972, 39(2): 131–152.
- [34] Gibb, B. C. 'Reproducibility'[J]. *Nature Chemistry*, 2014, 6(8): 653–654.
- [35] Rudin, C. 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'[J]. *Nature Machine Intelligence*, 2019, 1(5): 206–215.
- [36] Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., et al. 'Reproducibility in Machine-learning-based Research: Overview, Barriers, and Drivers'[J]. *AI Magazine*, 2025, 46(2): e70002.
- [37] Goldman, A., Beddor, B. 'Reliabilist Epistemology'[EB/OL]. *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/sum2021/entries/reliabilism/>. 2021–05–21.
- [38] Miller, K. 'Should AI Models Be Explainable? That Depends'[EB/OL]. *Stanford University Human-Centered Artificial Intelligence*, <https://hai.stanford.edu/news/should-ai-models-be-explainable-depends>. 2021–03–16.

[责任编辑 王巍 谭笑]