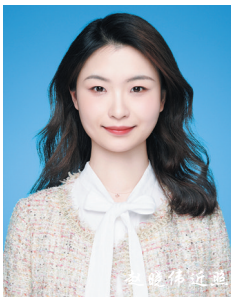
■ 教育数智化

DOI:10.15998/j.cnki.issn1673-8012.2025.01.002



人工智能赋能高校科学研究范式创新:价值、风险与进路

赵晓伟,王小雨,王艺蓉,沈书生
(南京师范大学　教育科学学院,南京　210097)

摘　要:高校是科学研究的主阵地,是基础研究与重大科技突破的主力军,肩负着加快形成新质生产力的重要使命。在新一轮科技革命浪潮下,人工智能技术在科学研究中的广泛应用与深度融合,不仅重塑了科学研究的基本过程,更引领着科学研究范式向智能科学研究范式转型。从逻辑实证主义视角出发,探讨人工智能赋能高校科学研究范式创新的多重价值,包括增强科学研究中的数据采集、加速科学假设的生成与验证、实现实验模拟与自动化,以及激发科学洞察与创意涌现,有助于理解 AI 赋能科学研究范式的价值和深远影响。基于此,审视当前科研范式转型过程中遭遇的诸多挑战与风险,包括人的边缘化风险("人在旁路"现象)、数据偏见与误导性主张的双重困境("数据陷阱"问题)、算法透明度的缺失("算法黑箱"问题)以及由此引发的信任危机等,高校在科学研究范式转型中应紧密依托新一轮科技变革,植根中国科学研究实践土壤,立足已有范式寻求创变,探索一条符合我国国情且具有鲜明中国特色的高校科学研究道路。具体来讲,应加强跨学科合作与协同创新、规范数据使用标准、推动可解释性 AI 的发展,建立健全科研伦理审查机制,为科研范式创新提供坚实的理论支撑与行动指南。

关键词:科学研究范式;智能科学研究范式;人工智能;数据陷阱;信任危机

[中图分类号]G644　[文献标志码]A　[文章编号]1673-8012(2025)01-0009-12

修回日期:2024-08-15
基金项目:国家社会科学基金重大项目"新一代人工智能对教育的影响研究"(VGA230012);认知智能国家重点实验室智能教育开放课题"新一代人工智能促进教育管理变革与政策创新"(iED2024-M002)
作者简介:赵晓伟,女,山东莒县人,南京师范大学教育科学学院讲师,博士,主要从事人机协同教育;
王小雨,女,江苏兴化人,南京师范大学教育科学学院硕士生,主要从事信息化教学设计;
王艺蓉,女,浙江杭州人,南京师范大学教育科学学院硕士生,主要从事教育技术研究。
通信作者:沈书生,男,江苏海安人,南京师范大学教育科学学院教授,博士生导师,主要从事教育技术理论研究。
引用格式:赵晓伟,王小雨,王艺蓉,等. 人工智能赋能高校科学研究范式创新:价值、风险与进路[J]. 重庆高教研究,2025,13(1):9-20.

Citation format: ZHAO Xiaowei, WANG Xiaoyu, WANG Yirong, et al. Empowering innovation in scientific research paradigms at higher education institutions with artificial intelligence: value, risks and approaches[J]. Chongqing higher education research,2025,13(1):9-20.

电子显微镜的发明使人类首次发现细胞内部的超微结构,大型强子对撞机的制造使希格斯玻色子的存在得以证实。不难发现,历次重大科学发现都离不开新工具或新技术的诞生。新一轮科技革命加速了科学研究的进程,从根本上改变了科学研究的方式,推动科学研究的范式转型。在数据密集型第四科学研究范式下,研究者依托庞大的数据集进行数据建模与分析,探索数据中的潜在规律。以生成式 AI 为代表的新一代 AI 技术,不仅能够高效完成大规模的数据计算,更能协助研究者发现潜在的科学问题,探寻新的研究方向,为科学研究提供新方法,推动科学研究向第五范式——智能科学研究范式(AI for Science,AI4S)转型。

AI4S 智能科学研究范式代表着 AI 数智技术与科学研究的深度融合和双向赋能[1],加速科学发现与研究的进程,实现科学研究过程的自动化和内容生产的智能化。与此同时,科学研究的发展推动着智能算法模型的持续优化与迭代。高校是科学研究的主阵地,高校研究人员是基础研究与重大科技突破的主力军。高校肩负着加快形成新质生产力、实现高水平科技自立自强、推动科学研究高质量发展的重要使命,应响应国家科技战略需要推进科学研究创新,以提升国家的原始科技创新能力。在新科技革命背景下,基础研究转化周期明显缩短,国际科技竞争形势激烈,人工智能驱动科学研究不仅跃升为政策制定者与战略规划者无法回避的核心议题,更是推动社会进步的关键。高校应以数智技术作为科学研究范式转型的主要引擎,推动人工智能与科学研究双向赋能,发挥人工智能在科学研究及其范式创新上的强大驱动力[2]。本研究梳理了 AI 赋能高校科学研究范式创新的价值与风险挑战,并针对性地提出应对策略,以期为高校科学研究范式的变革提供参考。

一、AI 赋能高校科学研究范式创新的价值意蕴

库恩指出,科学革命就是范式的更替,范式的确立与科学共同体的行为准则及其待解的谜题紧密相连,涉及某些实际科学实践的公认范例,为特定科学研究提供参照[3]。科学研究范式不仅是科学发现与科技创新的基石,更是科学共同体在特定时期进行科学研究的方式,契合科技创新的内在规律[4]。科学研究实质上是追寻“自然之谜”答案的过程,逻辑实证主义者认为这一过程遵循既定的科学探究模式。亨普尔(Hempel)将这一过程划分为 4 个阶段:观察记录事实、分析归类事实、归纳推导结论和检验结论[5]。科学发现被视为从提出假说至被接受的过程,Laudan 认为,此中间过程不仅是归纳,更需要在猜想或假设后进行求证与完善[6]。整体来看,科学发现大致包括观察提问、假说建立、实验验证、理论构建并持续修正的过程。

科学研究范式的演进经历了多个阶段,从实验描述的经验范式,到模型归纳的理论范式、仿真模拟的计算范式,再到数据密集型分析的数据范式,直到如今的智能科学研究范式(AI4S)。在新科学研究范式下,AI 增强了研究者的数据获取方式与计算能力,深刻改变了假设建立的逻辑,通过自动化或模拟实验过程,加速科学洞见的形成,展现了其强大的潜力。同时,AI 通过回溯、预见与迭代等方式,不断优化科学研究的过程,推动科学生产力的提升,基本重塑了科学研究的一般过程(如图 1)。

(一)助益科学研究中的数据采集

在科学研究领域,数据的采集转换与深度分析是构建科学见解和理论的基石,AI4S 背景下高校科学研究在数据采集与管理的各个环节已取得突破性进展。在数据发现与选择方面,深度自动编码器和其他先进的深度神经网络已展现出从庞大且结构复杂的研究数据中提取非线性特征、筛选高价值研究训练数据的能力。这些智能技术能够根据研究需求,自动识别和剔除异常值,获取更为纯净的数据,有效应对存储与传输大规模研究数据的技术挑战[7]。在数据标注方面,大型标签数据集对于模型的训练至关重要,随着标注策略的不断革新,自监督学习等先进方法已能够实现标签的自动标记,极大地减轻了研究者在数据标注上的负担,能够以更少的数据标签完成数据集的全面标注,为模

型训练提供有力支持。在数据生成方面,随着研究数据集质量与规模的不断提升,用于数据增强和模型训练的算法也持续迭代,不仅能够扩充数据集规模,还能集成不同类型的数据集,生成跨学科知识领域的数据,从而大幅提高数据集的综合性[8]。此外,AI 的发展还优化了数据细化处理的能力,提升数据质量的精度,为科学发现与洞见提供可靠的数据支撑。

图 1　AI 赋能高校科学研究范式的基本逻辑

芝加哥大学数据科学研究所 Willett 提出,科学家可以借助 AI 分析复杂的科学数据,如借助机器学习与深度学习技术,从韦伯望远镜捕获的宇宙图片中挖掘出潜在的高价值数据,从而揭示宇宙的奥秘[9]。在物理科学实验中,AI 能够辅助科学家对海量实验数据进行筛选、采集与存储。如在质子碰撞试验中,利用基于深度学习的自动编码器对碰撞产生的数据进行过滤性采集,能够有效应对数据存储的极限挑战[10]。算法模型的训练往往需要专业人员对庞大的数据集进行标注,而垂直领域模型的通用性和推广性往往受到限制,因此数据标注工作尤为繁重。伦敦大学学院和 Moorfields 眼科医院提出了一种名为 RETFound 的视网膜图像基础模型,该模型利用自监督学习,在超过 160 万张未标注的视网膜图像上进行训练,并在眼部疾病诊断、预后分析以及系统性疾病预测等任务中表现出卓越性能[11]。该成果不仅减轻了数据标注的负担,更展示了 AI 在科学研究领域的巨大潜力和广阔前景。

(二)加速科学发现中的假设生成

在科学发现的过程中,提出可供检验的研究假设是重要开端。传统科学研究范式下这一过程往往费时费力,深受研究者的专业知识、探索能力以及个人偏见的影响。AI4S 背景下,数智技术为假设生成方式带来革命性改变,提升了生成效率,推动科学见解的整合,加速推进隐式和未解的科学发现。假设生成方式主要包括两类:

一是基于文献的知识发现(literature-based discovery,LBD)。该方法运用文本挖掘、自然语言处理及机器学习等技术为模型提供测试数据集,对海量文献数据进行深度处理,经训练的模型能够自动从文献中抽取知识概念及隐含关联,助力研究者高效提出研究假设。目前,LBD 在生物医学、材料科学等领域已取得显著成效,预计下一代模型将具备处理图表、代码等多模态研究数据能力。例如,劳伦斯伯克利国家实验室 Tshitoyan 团队采用自然语言处理的无监督学习 AI 技术,分析材料科学领域的论文摘要,发现训练后的系统具备“化学直觉”,能够预测具有特定属性的新材料,并提出可能材料的假设与预测[12]。芝加哥大学社会学家 Sourati 等人进一步扩展了该方法,并将论文作者间的关联纳入考虑,训练了兼顾论文摘要与作者的 LBD 系统,发现该系统预测材料的能力达到 Tshitoyan 团队的

两倍[13]。

二是智能辅助的假设空间聚焦[8]。主要包括 3 个过程:首先,假设对象的快速筛选。科学研究中一个科学问题往往包含多个可能的假设对象或符号,构成假设空间时并非所有假设都值得深入探索,AI 将辅助科研人员快速筛选对象或符号,聚合最有可能产生价值的假设范围。例如,利用自监督学习的 AI 预测器,可以根据已有知识和数据,预测哪些对象或符号更有可能产生有意义的假设。其次,提升假设的质量和生成效率。一旦确定要探索的假设空间,就需要生成具体的假设,AI 可以通过优化算法和强化学习技术,助力生成高质量的假设。具体而言,可训练 AI 代理(如强化学习模型),在假设空间中寻找高质量的候选假设,该代理将根据需要采用“奖励策略”的方法,不断尝试和改进,直到找到满足要求的假设。最后,实验前的假设评估与预测。在确定待探索的假设后,还需要对其加以评估和预测,以确定哪些假设最值得进行实验验证。该过程同样可以借助 AI 的支持,通过整合先前的知识和数据,学习研究假设的贝叶斯后验分布(即假设在给定数据下的概率分布),可使用预测模型对假设进行实验前的对比和评估,选择最有可能成立的假设进行实验。该方法可以帮助研究者更准确地预测实验结果,减少实验浪费,提高科学研究效率。

(三)实现科学实验模拟与自动化

实验作为检验科学假设的关键,是推动科学发现的核心环节。传统实验面临成本高昂、效率低下、精度不足、难以实施等诸多困境,限制了科学研究的步伐。AI4S 背景下科学实验迎来模拟化与自动化的革命,不仅降低实验成本、提高效率,更以空前的规模自动化开展,拓展科学实验边界的同时,更在深层次上重塑科学实验的本质。科学实验模拟化即利用计算机模拟科学实验的全过程,得益于高性能计算机与深度学习技术的进步,大模型、大数据与大算力的融合,科学实验朝着计算机模拟的方向飞速发展,为复杂实验的实施开辟了新途径,目前已在多领域取得显著成效。如在气象科学领域,DeepMind、华为和 Nvidia 等科技公司推出了 AI 天气预测模型,不仅提高预报的速度和准确度,更降低了预测成本。Nvidia 公司更是计划通过“Earth-2”项目开展更大区域的气候预测,以展现 AI 在气象科学中的巨大潜力[14]。在化学领域,麻省理工学院研究人员通过机器学习模型成果模拟了化学反应中的过渡态这一短暂过程,加速了化学反应物和催化剂的研究进程[15]。

科学实验自动化的实现则依托大语言模型、人工智能、机器人等先进技术的支持。在该模式下,大型科学研究平台或自动实验系统能够独立设计实验并执行,缓解人类实验员的压力。从实验前的方案制定、方法选择和环境条件配置,到实验中的数据记录、参数监控与调整,再到实验后的数据分析与结论生成,AI 均能胜任。这一变革极大地解放了科学研究的生产力,科学家们能够以前所未有的速度完成实验。例如,密歇根大学研发的 BacterAI 平台,能够实现细菌实验的高度自动化,每日可完成上万次实验,全程无需人工介入[16]。中国科学技术大学研发的机器人化学家“小来”,由“化学大脑”、机器人实验员和化学工作站组成,其中“化学大脑”能阅读文献、设计实验、优化方案,机器人实验员与工作站协同执行实验,在短短 5 周内完成传统方法需耗时 1 400 年的化学实验[17]。自动化实验的另一优势在于高效执行重复性实验,科学界普遍存在对重复他人工作的动力不足的现象,自动化实验机器人能够对先前研究成果进行重复性实验,节省研究者的精力,不仅验证已有科学主张,还详细记录实验步骤,为后续学术分析提供宝贵数据,推动科学研究的可持续发展。例如,曼彻斯特大学 Roper 教授团队借助 Eve 机器人成功复制多项生物医学研究成果,为科学发展注入新活力[18]。

(四)激发科学预测中的洞见涌现

产生科学知识是科学研究的价值所在,而科学预测作为这一过程中的关键一环,其准确性直接决定着科学知识产生的步伐。传统科学预测往往依赖科学家的理性分析、演绎推理与抽象建模。AI4S 背景下科学预测方式正经历深刻变革,通过预训练模型能够实现无需人类干预的科学预测,这一转变

不仅重构科学预测的实践路径,更为多个领域催生丰富的科学洞见。基于 AI 的科学预测,融合深度学习算法与庞大数据集的训练,结合学科领域的先验知识,构建高度专业化和智能化的预测模型。这些模型具备预测新物质、新结构或新变化的能力,能为研究者提供全新的研究视角与方向。同时,模型强大的特征分析与模式识别能力,能够高效处理大规模、高维度数据,实现自动化的数据处理与结果预测,极大提高预测效率与精度,减轻研究者的负担,提升科学研究生产力。更重要的是,这种预测方式有助于打破科研中个体的思维框架与限制,促进多学科交叉和跨学科的知识共享,激发新的研究思路与方法,助力集体智慧的涌现,推动科学发现边界的不断拓展。

一些研究案例证明了 AI 在科学预测中的巨大潜力和广阔前景。麻省理工学院的科学家利用经过约 2 500 个分子训练的 AI 模型,成功从数百万种候选化合物中预测出能够对抗“超级细菌”的新型抗生素 halicin,并通过生物实验验证了其有效性[19]。在生物分析领域,AlphaFold2 蛋白质结构解析模型能够迅速预测目标蛋白质的结构,甚至揭示自然界中尚未发现的蛋白质结构。此外,美国加州大学伯克利分校和劳伦斯国家实验室的“A-Lab”自主实验系统,展现出强大的自主发现和预测新材料的能力。当前,大模型正逐步向多模态通用性发展,研究者和开发者能够更便捷地训练自定义模型,满足不同领域科学研究的需求。加州理工学院、麻省理工学院及 Nvidia 公司,基于开源大语言模型联合构建了定理证明器,为数学领域的形式化定理证明提供了有力支持[20]。微软研究院研发的 DiG 模型框架,在处理不同类型的分子系统和描述符方面表现出色,主要用于预测分子结构平衡分布,为分子科学研究、药物设计以及材料科学等领域的研究带来新机遇[21]。IBM RXN 云服务则基于数百万个合成有机化学反应训练的 Transformer 模型,为化学反应预测和有机合成实验提供新思路。

二、AI 赋能高校科学研究范式创新的风险检视

数智技术无疑为高校科学研究注入了前所未有的活力,显著加速科学研究进程,拓宽科学发现视野。然而,AI 的双刃剑效应也随之显现,新技术浪潮下亟须审视 AI 引发的系列风险挑战。对高校科学研究中存在的主体工具性依赖、数据偏见导致的误导性主张、算法的复杂性与不透明性引发的解释困境,以及虚假内容生成带来的信任危机等问题,必须予以高度警觉和审视剖析。

(一)人在“旁路”:技术依赖导致主体性缺失

AI4S 背景下,数智技术与科学研究的深度融合无疑增强了研究的客观性和严谨性。然而,这种深度融合也伴随着一种隐忧,即过度依赖 AI 技术可能导致研究者的主体性地位逐渐淡化。

一方面,研究者期待通过 AI 消除科学研究过程中难以避免的主观性、偏见与认知局限,以实现科学设计与实施的优化。然而,当这种期待转变为对 AI 的过度依赖,甚至将其视为“万能钥匙”时,人的主体性在科学研究中便逐渐丧失,从而引发一种被称为“理解幻视”的认知风险。所谓“理解幻视”,即主体错误地认为借助 AI 技术能弥补自身的认知局限,进而产生过度的自信,认为自身对研究对象的了解已经超越实际水平。这种过度依赖不仅阻碍研究者深入研究和广泛探索,更可能限制科学研究的发展空间。

另一方面,大型科学研究平台与自动化实验系统的广泛应用,确实显著提升了实验的效率,加速了科学发现的步伐。然而,在高度自动化的实验过程中,人的决策性地位逐渐削弱,不再占据科学研究的主导地位。AI 研究系统能够自动完成文献搜索、综述、研究方向生成、研究假设提出,甚至提供完整的研究设计方案,并自动化处理庞大数据集以产生研究结论。AI 在提升科学研究效率的同时,也引发科学研究本质的变化。在该模式下,AI 似乎成为科学知识创造的主体,研究者则似乎被置于科学研究的“旁路”。在此状况下,主体的批判性思维面临严峻挑战,科学创造被限于 AI 自主生成的孤岛中,导致科学创意匮乏。缺乏人类批判性思维的参与,科学研究往往难以触及核心问题,也难以

产生真正的影响力,这也是新技术背景下科学研究成果在数量上显著增长但具有颠覆性价值的研究却逐渐减少的可能原因之一[22]。诚然,数智技术在科学研究中发挥着不可或缺的作用,但它始终只能作为得力工具,人类认识与改造世界仍然离不开头脑中的“奇妙计算机”。

(二)数据陷阱:数据偏见引发误导性主张

数据作为对事物客观描述的基石,若在收集、处理与分析的各个环节中遭遇边缘化、覆盖面不全或样本偏向性等问题,基于数据生成的结果便会陷入错误性、不完整性的境地,进而引发误导性主张危机。AI 模型训练过程的复杂性与阶段性使得数据偏见显现于多个环节。

在数据采集阶段,用于模型训练所采集的数据,其来源的局限性、种类的单一性以及标签划分的主观性,都有可能引发数据偏见。耶鲁大学的一项研究揭示了这一点,利用 AI 的皮肤病学算法在诊断不同种族人群时,准确性存在显著差异,这背后的原因,正是训练模型所采集的数据大部分来自白人,导致算法应用在其他种族时产生偏见[23]。

在模型训练阶段,训练过程的不严谨同样会埋下偏见的隐患。例如,训练数据和测试数据集划分不明确、充斥重复数据、混杂伪数据,都可能导致算法模型本身带有偏见。普林斯顿大学 Kapoor 团队指出,至少在 17 个领域和数百篇论文中存在可重复性危机,这背后正是基于伪研究数据所训练的模型,其生成的研究结果也具有误导性[24]。同时,模型训练集和测试集的混淆,也将引发拟合、功能泛化、能力不足等问题,进一步加剧科学研究的质量危机,并引发 AI 知识生成的可信度危机。

开发者的个人偏向也将对算法产生影响。由于开发者基于自身的认知偏见与情感倾向,在算法设计时可能会不自觉地强调某一特定特征而忽视其他关键变量,这种偏见性在算法的实际应用中表现为对不同用户或系统的不公平。同时,算法设计群体存在单一性特征。美国平等就业机会委员会观察到“高科技”行业中白人比例过高,这一现状将加剧算法开发中的种族歧视,使模型无形中偏向部分特定群体,并对其他群体产生歧视[25]。

(三)算法黑箱:不透明决策导致解释困境

科学研究中的算法“黑箱”现象日益凸显,这源于技术本身的复杂性与技术开发者的算法保护策略。研究者在使用 AI 时,往往无法洞察算法生成结果背后的设计逻辑与内部机制,导致对算法结果的解释、监管与审查陷入重重困境。这种不透明性不仅影响研究结论的可靠性,也阻碍研究者对算法决策的信任。一方面,研究中算法与模型设计的科学性不足,加之算法内部复杂的数据和参数设置,使得算法的运作过程和决策逻辑难以被完全理解,研究者在使用这些模型时,可能会基于不准确的预测或误导性的结论开展研究,进而对整个人类知识库构成威胁。例如,研究人员在调整数据和参数以符合个人预期时,AI 为其提供了过度的自由度,可能导致研究结果失真。另一方面,算法的不透明性也限制了研究者难以准确追溯问题根源,不知不觉中可能陷入学术信息茧房,导致视野受限,创新受阻。此外,推荐算法机制倾向于推送与用户偏好高度一致的内容,将加剧茧房效应,限制研究者的研究视角与深度。

算法的复杂性和不透明性也将加剧科学研究内容审查的困难。一方面,由于无法直接窥视算法的内部运作,研究者往往难以对算法生成的结果进行有效评估与监管,从而难以确保研究结论的准确性和可信度。另一方面,科学研究活动中涉及数据收集、处理、分析和解释等环节,算法的不透明性会使得研究者难以确保项目的合规性和伦理性,引发数据滥用、隐私泄露等伦理问题。此外,法律法规、伦理标准以及监管机构的审查方式具有一定滞后性,无法及时应对算法带来的各种风险。以生物医学领域为例,AI 算法在图像分类中的应用虽然具有巨大潜力,但错误的分类结果可能直接关系到患者的生死。因此,确保算法的透明性和可靠性,是科学研究领域亟待解决的问题。

(四)信任危机:大模型滋生虚假内容泛滥

AI 特别是生成式 AI 的迅猛发展,使其内容生成能力达到了前所未有的高度。然而,这种技术的

滥用却导致学术造假问题频发。尤其在高校中,学生直接利用生成式 AI 技术完成课程论文和作业,引发严重的学术诚信危机。马里兰大学的学生行为办公室在 2022—2023 学年中收到的与 AI 相关的学术诚信违规举报高达 73 起,令人震惊的数字背后,凸显高校学术诚信问题的紧迫性[26]。在科学研究领域,生成式 AI 的广泛应用与论文发表所要求的原创性声明原则之间存在明显的冲突。研究表明,当 ChatGPT 创建的摘要被提交给学术评审员时,仅有 63% 的“赝品”被发现[27]。这一发现揭示了学术造假监管工作的艰巨性,也让我们意识到,在享受技术带来便利的同时,必须警惕其潜在的风险。确保科学研究的学术公正性,对于维护科学研究生态的稳定运行至关重要。高校作为学术培养和知识生产的主要场所,理应承担起监管和治理的重任。

AI 大语言模型在科学研究中的广泛应用确实带来诸多便利,但频繁生成的虚假信息以及缺乏有效监管机制的现实,已经引发研究者对 AI 技术的信任危机。一方面,模型的训练依赖庞大的训练集,然而训练集来源的多样性、复杂性与冗杂性,使得数据的真伪性与准确性难以保证。此外,人为的恶意训练与对抗攻击,也可能导致模型生成虚假、错误或误导性内容。另一方面,相较于传统实验研究,利用 AI 进行实验与研究往往缺少必要的申请、批准环节,以及缺乏如 IRB(机构审查委员会)机构的督查。科学研究中 AI 使用的规范与伦理尚未得到系统、标准的管理,其使用过程缺乏透明度和可控性,进一步加剧研究者对人工智能的信任危机。*Nature* 对 1 600 名研究人员的调查显示,超过半数的研究者担心 AI 生成的错误内容激增,认为抄袭变得更容易且难以检测,担忧其可能导致错误或不准确的研究结果,更有近六成的研究者认为这些工具可能引发欺诈行为[28]。这一担忧不容忽视。

三、AI 赋能高校科学研究范式创新的实践进路

科学研究范式的转型并非意味着对过往范式的彻底背离,而是要在已有范式的基础上持续创变。高校作为实现高水平科技自立自强的关键力量,应积极融入数智技术浪潮,利用技术变革为科学研究创新注入新动力。同时,需保持敏锐的洞察力,主动应对技术衍生的风险与挑战,在实践中探寻一条深植于中国学术土壤且充分展现中国独特话语体系的高校科学研究发展道路,以推动科学研究的持续繁荣与进步。

(一)人在回路:以跨界协同提升 AI 科研素养

高校在开展科学研究的过程中,应始终坚守人的主体地位,避免陷入“理解幻视”困境。所谓“人在回路”,指机器决策中强调人的决策参与,构建融合人类智慧的算法,通过跨界融合提升研究者的 AI 科研素养,并设计相应的培训方案,确保“科学研究—人才培养”良性循环和可持续发展。

其一,开发以人为本的智能科研系统,构建人在回路的 AI 算法。这一系统强调人机协同,其中人的作用与地位至关重要,AI 作为辅助者而非替代者。系统需要具备高度的可干预性和适应性,使研究者能随时干预和调整系统运行,确保科学研究方向的正确性。同时,算法的设计应具备透明性、可解释性,确保研究者能够清晰理解算法的运行过程和结果,从而进行准确的评估。以华盛顿大学研究团队为例,他们在人机协同科学研究过程中,充分发挥人的主导作用,结合 AI 技术提议,采用扩展 RFdiffusion 与序列设计工具 ProteinMPNN 结合的新方法,充分发挥了实验设计中人类思维的创新性[29]。

其二,倡导各界紧密合作,跨界融合培育 AI 科研素养。通过建立“高校—政府—企业”三位一体的跨界协作平台,构建“AI+X”跨专业融合课程体系,推动 AI 技术与其他学科领域的深度融合,全面加强研究者 AI 科研素养的培养[30]。譬如,麻省理工学院联合 8 个不同院系,依托“计算机科学与 AI 实验室”,为不同需求层次的学生提供多样化的学位项目(包括计算与认知工程硕士、计算机科学博士、计算机科学与工程博士等),并积极与政府、学术研究机构等合作,为学生提供丰富的实习机会[31]。帝国理工学院联合康奈尔大学以及政府、行业,建立了跨大西洋 AI 科学网络,发布 I-X 创新

计划,将关键学科(如材料、物理学、生物学或可持续科学领域)的 AI 科学家聚集起来,共同探索该技术在科学工程研究中的应用,以加速科学发现[32]。

其三,依托跨界共同体,构建学分认定与资金保障机制,为跨学科融合培育提供支持。在学分互认方面,可以为科学研究后备军提供灵活多样的学习路径和成果认证方式。如我国华东地区 6 所高校联合华为、百度、商汤科技等企业,采用共建共选的方式,以小规模限制性在线课程(SPOC)的形式,开设了“AI+X 微专业”,学生在 1-2 年内获得不少于 12 个学分,就可被授予由 6 所高校共同签章的微专业证书,学分可以申请转换为校内公共选修课学分[33]。在资金支持与激励方面,Schmidt Futures 项目面向高校设立了高达 1.48 亿美元的 AI 科学奖学金,培养未来的 AI 领导者,目前已在促进数学发现、监测作物干旱、改进太阳能材料等多个领域取得显著成效[32]。

(二)数据规范:以数据标准确保合规使用

AI 模型的训练和学习过程离不开数据的反复训练与测试,数据的客观性、中立性、全面性,数据库的标准化程度,都将影响 AI 算法决策的结果。若数据集在划分过程中缺乏严谨性,还可能导致模型偏见问题。因此,高校科学研究应致力于数据规范化,构建清晰的数据清单和统一的数据标准框架。

首先,明确数据标准,建立详尽数据清单。通过系统化数据存储与精细化管理,确保数据的准确性、完整性和一致性。同时,遵循统一的标准进行数据的标准化操作,明确数据的来源、种类和标签的划分。目前,FAIR 标准(即可查找、可访问、可互操作和可重复使用)已在多个领域得到广泛认可,如跨学科地球数据联盟 EarthChem 图书馆,围绕业界标准整合数据集格式,与不同学科领域保持高度一致,并能在互联网上便捷地获取[34]28。

其次,基于标准化数据,构建高质量数据集和数据库。建议从国家层面推动跨行业、高校、企业以及科学研究机构的合作,共同构建超大型数据集和数据库,为 AI 技术的发展提供坚实的数据支撑。在国际方面,谷歌联合哈佛大学、麻省理工学院等高校,依据统一数据标准整合处理了包含 25 000 TB 小鼠大脑特定区域图谱数据集,为相关研究提供宝贵数据源[35]。在国内,腾讯优图实验室联合多个部门,基于统一标准建立了全球最大的甲骨文单字数据库,为后续相关研究工具的开发提供支持[36]。

再次,明晰数据使用标准,规范模型训练过程。模型训练过程中操作不当,将导致数据偏见等问题,进而影响 AI 模型的准确性。因此,需要建立相对统一的标准,包括数据使用、模型训练策略选择、评估与反馈指标等。国内外众多机构已发布数据使用标准和指南,如 FAIR 原则中对数据在代码中的存放、研究论文的引用进行了规范[34]26;欧盟委员会 FAIR 数据专家组的报告对数据的引用所需步骤进行了罗列与说明[34]26;澳大利亚研究数据共享组织(ARDC)制作了在线培训指南,引导研究人员正确引用数据、样本和软件资源[37];美国研究机构发布了如何增加科学数据访问权限的指南,以规范数据的使用[38];斯坦福大学、圣地亚哥州立大学等高校发布相关生成式 AI 教学指南与使用指南,分别对教师、学生等不同群体提供使用手册,并围绕使用的过程步骤进行指导与规范[39]。

(三)可解释 AI:以透明算法研发解密“黑箱”

算法模型在数据处理中直接输出结果,其内部运行机制复杂,即使对算法工程师或数据科学家而言也是一片迷雾,这种不透明性极大地削弱了决策者对 AI 技术的信任。为应对“黑箱”挑战,可解释 AI(XAI)应运而生。XAI 系统致力于揭示 AI 模型的内部逻辑,详细解释 AI 决策和预测的依据与过程,揭示潜在的影响和偏差,并展现模型的准确性、公平性、透明度和结果。其不仅为用户提供理解和信任算法结果的途径,也支持技术人员以负责任的态度进行开发,同时赋予决策者质疑 AI 结果并据此作出明智决策的能力。在推动 AI 变革科学研究范式的过程中,高校科学研究应积极采用 XAI 工具,以提高算法模型的透明度,帮助研究者和 AI 开发者更好地理解和信任新一代的 AI 科研伙伴。

在高校科学研究领域,XAI 的价值已得到广泛认可。加州理工学院在虚拟研讨会上,明确将开发

XAI 列为重要议题,强调 AI 模型应具备精确的开发校准、可验证的单元测试以及用户友好的界面,确保实际应用中的高效性和可靠性[40];UCLA 的研究团队利用 XAI 技术对原本用于山体滑坡预测的深度神经网络(DNN)进行优化,成功开发出 SNN 模型[41],不仅改善了 DNN 在解释性方面的不足,还保持了高准确性、高泛化能力和低模型复杂度,为山体滑坡的精准预测提供有力支持。在国内,山东大学开发的 RetroExplainer 算法在有机物的逆合成路线识别中取得显著进展,不仅提高了识别效率,还为用户提供了对算法决策过程的深入理解[42];清华大学的研究团队利用可解释的机器学习技术对 BiVO4 光阳极系统进行优化,利用 AI 模型分析过去的实验数据,揭示了系统内部的复杂关系,为进一步优化提供重要指导[43]。目前,在科学研究中用于缓解算法"黑箱"问题的 XAI 主要有两类实现路径。一类是设计本质上可解释的预测模型,这类模型通过采用基于规则、决策树和线性模型等特征权重、内部路径和规则可见的组件,确保决策过程的可追溯性和透明度[44]。然而这种方法也面临功能和预测性上的限制,因为基于规则的模型组件往往较为固定,用户只能进行微调。另一类是事后可解释性技术,主要解决深度学习模型在追求预测准确性时牺牲透明度和可解释性的问题[45],包括全局模型和局部模型两个维度,前者揭示黑盒模型的平均行为,后者则针对单个预测进行解释。

(四)伦理审查:以科研治理促进可持续发展

AI 赋能科学研究的同时,引发了虚假信息生成和学术造假问题,这导致科学研究领域对 AI 的信任度下降。因此,对 AI 进行伦理审查与治理已成为当务之急。

首先,积极开发用于识别造假的 AI 工具,利用 AI 技术对造假问题及虚假信息进行规范与治理,并不断提升治理效率和准确性。政府应给予技术创新研究资金支持及资源激励,确保技术的可持续发展。例如,OpenAI 的 Deepfake 探测器能有效检测 DALL-E 等图像生成器创建的虚假内容[46];百度 AI 开放平台推出了图片造假审查工具,能够精准识别图片是否经过篡改;哈佛大学在医疗 AI 领域投入巨资,集结顶尖科学研究人才,实现了技术创新突破[47]。

其次,设立大模型伦理委员会,统筹各部门资源共同应对 AI 伦理风险。应汇聚各行业领域的共治力量,形成治理共同体,并重视社会中每个成员的力量,使其具有对技术限制的发言权。正如比勒陀利亚大学数据科学主席 Marivate 所言,AI 治理是一项团队活动,道德决策和责任应由全社会共同承担[48]。

再次,尝试构建全面 AI 科学研究治理框架,阐明治理体系、机制和条例,鼓励公众参与治理,提供具体参与路径,通过培训研讨、课程开发等途径,推广 AI 研究工具,普及应用伦理规范,制定 AI 研究工具的使用原则、规定与指南,从而推动治理流程的高效运转[49]。在框架明晰的基础上,应进一步细化实施策略,例如设立专门的治理机构,严格监测大模型输出,确保输出内容的真实性与准确性;加强算法的备案和变更管理,对算法的合法性、合规性和安全性进行全面审查与评估。此外,政府、科研机构、企业等各方应密切合作,共同明确治理流程、框架及实施路径。英国政府在数据伦理方面树立了典范,不仅制定了详细的数据伦理框架(涵盖内部监测与治理机制、数据处理责任、数据处理活动记录等方面),还明确了数据处理中的责任,提出公众问责制,确保公众或其代表能够有效监督政府决策与行动[50]。

四、结　　语

从寻找数据模式的机器学习模型,到从海量文本代码中生成内容的最新通用算法,人工智能不仅加速了知识探索的步伐,更为科学探索开辟了新航道。在人工智能赋能的新科学研究工作模式下,科学研究工作者的想象力不断被激发,能够洞察潜在的数据模式与异常现象,跨越自身的认知边界,对已有研究作出新解释,并发现更具价值的新问题。人工智能赋能科学研究改变了工作的流程,引领科学研究范式的创变。新科学研究范式中人工智能引领知识发现,成为人类创造力的强大催化剂,依托数据驱动

的预测模型,融合模拟技术与可扩展计算的力量,激发了研究者的创新思维,形成创新性的解决方案。

新局面意味着新风险,高校作为科学研究的主阵地,必须以高度的责任感与前瞻性的战略眼光进行全局性部署,确保人工智能在科学研究中可靠运行,同时采取周密的伦理审查措施。如此,才能实现人工智能益处最大化、潜在风险最小化,解锁以前无法挖掘的科学奥秘,推动人类社会迈向更加辉煌的未来。展望未来,人工智能与科学研究的深度融合将引领一系列前沿方向,涵盖从方法论的革新到自主科学研究系统的构建,再到知识提取、复杂现象建模、假设生成验证及新学科建设与跨学科合作的全方位推进,在提升数据质量、优化实验设计、挖掘潜在模式等方面发挥关键作用,在新兴科学领域人工智能更进一步展现巨大潜力,为科学研究开辟新道路。

参考文献:

[1] 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.

[2] 张海生.人工智能赋能学科建设:解释模型与逻辑解构[J].高校教育管理,2023,17(3):42-50,75.

[3] 库恩.科学革命的结构[M].金吾伦,胡新和,译.北京:北京大学出版社,2003:9.

[4] 张海生.人工智能赋能大学治理:多重效应与治理效能转化[J].重庆高教研究,2024,12(2):25-36.

[5] 亨普尔.自然科学的哲学[M].张华夏,余谋昌,鲁旭东,译.北京:中国人民大学出版社,1986:12.

[6] 周林东.科学哲学[M].上海:复旦大学出版社,2004:128.

[7] ZHOU C, PAFFENROTH R C. Anomaly detection with robust deep autoencoders[C]//Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. New York: Association for Computing Machinery,2017:665-674.

[8] WANG H C, FU T F, DU Y Q, et al. Scientific discovery in the age of artificial intelligence[J]. Nature,2023,620(7972):47-60.

[9] How AI is transforming scientific research, with Rebecca Willett(EP. 117)[EB/OL].(2023-08-10)[2024-06-18].https://news.uchicago.edu/how-ai-transforming-scientific-research.

[10] GOVORKOVA E, PULJAK E, AARRESTAD T, et al. Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider[J]. Nature machine intelligence,2022,4(2):154-161.

[11] ZHOU Y K, CHIA M A, WAGNER S K, et al. A foundation model for generalizable disease detection from retinal images[J]. Nature,2023,622(7981):156-163.

[12] TSHITOYAN V, DAGDELEN J, WESTON L, et al. Unsupervised word embeddings capture latent knowledge from materials science literature[J]. Nature,2019,571(7763):95-98.

[13] SOURATI J, EVANS J A. Accelerating science with human-aware artificial intelligence[J]. Nature human behaviour,2023,7(10):1682-1696.

[14] REMI L, ALVARO S G, MATTHEW W,et al. Learning skillful medium-range global weather forecasting[J]. Science,2023,382(6677):1416-1421.

[15] DUAN C R, DU Y Q, JIA H J, et al. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model[J]. Nature computational science,2023,3(12):1045-1055.

[16] COLE A. Four Ways AI Has Already Changed Science[EB/OL].(2023-07-18)[2024-06-18].https://www.techopedia.com/four-ways-ai-has-already-changed-science.

[17] 王敏."机器化学家"是怎样炼成的[N].中国科学报,2023-05-16(01).

[18] The Economist. Could AI transform science itself?[EB/OL].(2023-09-13)[2024-06-18].https://www.economist.com/science-and-technology/2023/09/13/could-ai-transform-science-itself.

[19] STOKES J M, YANG K, SWANSON K, et al. A deep learning approach to antibiotic discovery[J]. Cell, 2020,181(2):475-483.

[20] YANG K,SWOPE A M,GU A,et al. Leandojo:theorem proving with retrieval-augmented language models[EB/OL].(2023-10-16)[2024-06-18].https://arxiv.org/abs/2306.15626.

[21] ZHENG S X, HE J Y, LIU C, et al. Predicting equilibrium distributions for molecular systems with deep learning[J].

Nature machine intelligence, 2024,6(5):558-567.
[22] PARK M, LEAHEY E, FUNK R J. Papers and patents are becoming less disruptive over time[J]. Nature,2023,613(7942):138-144.
[23] ZACK T, LEHMEN E, SUZGUN M, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study[J]. The Lancet digital health,2024,6(1):e12-e22.
[24] Nature. AI will transform science-now researchers must tame it [EB/OL]. (2023-09-27) [2024-06-18]. https://www.nature.com/articles/d41586-023-02988-6.
[25] SOLOMON D, MAXWELL C, CASTRO A. Systematic Inequality and Economic Opportunity[EB/OL]. (2019-08-07) [2024-06-18]. https://www.americanprogress.org/article/systematic-inequality-economic-opportunity/.
[26] GAUR A. UMD records at least 50 AI-related academic integrity cases in 2022-23[EB/OL]. (2023-11-27) [2024-06-18]. https://dbknews.com/2023/11/27/umd-chatgpt-academic-integrity-cases/.
[27] HOLDEN T H . ChatGPT is fun, but not an author[J]. Science,2023,379(6630):313.
[28] VAN N R, PERKEL J M. AI and science: what 1,600 researchers think[J]. Nature,2023,621(7980):672-675.
[29] RUMFIEL G. New proteins, better batteries: Scientists are using AI to speed up discoveries[EB/OL]. (2023-10-12) [2024-06-18]. https://www.npr.org/sections/health-shots/2023/10/12/1205201928/artificial-intelligence-ai-scientific-discoveries-proteins-drugs-solar.
[30] 祝智庭,戴岭,赵晓伟,等.新质人才培养:数智时代教育的新使命[J].电化教育研究,2024,45(1):5-260.
[31] 李锋亮,庞雅然.世界一流大学如何建设人工智能学科[N].光明日报,2022-02-10(14).
[32] JOHNS S. Imperial and Cornell University to cooperate on AI in scientific discovery[EB/OL]. (2023-04-18) [2024-06-18]. https://www.imperial.ac.uk/news/244392/imperial-cornell-university-cooperate-ai-scientific/.
[33] 吴飞,陈为,孙凌云,等.以知识点为中心建设 AI+X 微专业[J].科教发展研究,2023,3(1):96-116.
[34] European Commission, Directorate-General for Research and Innovation. Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data[R]. Luxembourg: Publications Office of the European Union,2018.
[35] JANUSZEWSKI M. Google Research embarks on effort to map a mouse brain[EB/OL]. (2023-09-26) [2024-06-18]. https://research.google/blog/google-research-embarks-on-effort-to-map-a-mouse-brain/.
[36] 腾讯 AI 新成果:建立全球最大的甲骨文单字数据库 [EB/OL]. (2023-07-25) [2024-06-18]. https://cloud.tencent.com/developer/article/2305113.
[37] ARDC. Citation and Identifiers [EB/OL]. (2022-05-14) [2024-06-18]. https://ardc.edu.au/resource/citation-and-identifiers/.
[38] HOLDREN J P. Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research[EB/OL]. (2013-02-22) [2024-06-18]. https://rosap.ntl.bts.gov/view/dot/34953.
[39] Artificial Intelligence Teaching Guide[EB/OL]. (2024-06-18) [2024-06-18]. https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide.
[40] Caltech Explainable AI Virtual Workshop[EB/OL]. (2021-09-23) [2024-06-18]. https://sites.astro.caltech.edu/xai4s/program.html.
[41] UCLA. UCLA team uses artificial intelligence to predict landslides-UCLA[EB/OL]. (2024-06-18) [2024-06-29]. https://www.chemistry.ucla.edu/news/ucla-team-uses-artificial-intelligence-to-predict-landslides/.
[42] WANG Y, PANG C, WANG Y Z, et al. Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks[J]. Nature communications,2023,14(1):6155.
[43] HUANG M, WANG S, ZHU H. A comprehensive machine learning strategy for designing high-performance photoanode catalysts[J]. Journal of materials chemistry, A. materials for energy and sustainability,2023,11(40):21619-21627.
[44] RAI A. Explainable ai:from black box to glass box[J]. Journal of the academy of marketing science:official publication of the academy of marketing science,2020,48(1):137-141.
[45] MARGOT V. A Brief Overview of Methods to Explain AI (XAI)[EB/OL]. (2021-11-27) [2024-06-18]. https://towardsdatascience.com/a-brief-overview-of-methods-to-explain-ai-xai-fe0d2a7b05d6.
[46] METZ C, HSU T. OpenAI Releases "Deepfake" Detector to Disinformation Researchers[EB/OL]. (2024-05-07)

[20240618]. https://www.nytimes.com/2024/05/07/technology/openai-deepfake-detector.html.
[47] Artificial Intelligence in Medicine Program[EB/OL].[20240618]. https://aim.hms.harvard.edu/.
[48] FRUEH S. How AI Is Shaping Scientific Discovery[EB/OL].(20231106)[20240618]. https://www.nationalacademies.org/news/2023/11/how-ai-is-shaping-scientific-discovery.
[49] 常桐善,赵蕾.美国高校应对和使用人工智能工具的策略与原则[J].重庆高教研究,2024,12(4):68-79.
[50] Central Digital and Data Office. Data Ethics Framework[EB/OL].(20200916)[20240618]. https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020.

(责任编辑:杨慷慨　校对:吴朝平)

Empowering Innovation in Scientific Research Paradigms at Higher Education Institutions with Artificial Intelligence: Value, Risks and Approaches

ZHAO Xiaowei, WANG Xiaoyu, WANG Yirong, SHEN Shusheng
(*College of Educational Sciences*, *Nanjing Normal University*, *Nanjing* 210097, *China*)

Abstract: Colleges and universities are the main battlefield of scientific research. As the main force of basic research and major scientific and technological breakthroughs, university researchers shoulder the important mission of accelerating the formation of new quality productivity. Under the wave of the new round of scientific and technological revolution, artificial intelligence technology is widely applied and deeply integrated in various aspects of scientific research, not only reshaping the basic process of scientific research, but also leading the transformation of scientific research paradigm to the intelligent science research paradigm. From the perspective of logical positivism, exploring the multiple values of AI empowering innovation in scientific research paradigms in universities, including enhancing data collection in scientific research, accelerating the generation and verification of scientific hypotheses, achieving experimental simulation and automation, and stimulating scientific insights and creativity, can help understand the value and profound impact of AI empowering scientific research paradigms. Based on this, the challenges and risks encountered in the current transformation of scientific research paradigms were exmined, including the risk of human marginalization (the phenomenon of "human-on-the-side"), the dual dilemma of data bias and misleading claims (the "data trap" issue), the lack of algorithm transparency (the "algorithm black box" issue), and the resulting crisis of trust. Universities should closely rely on the new round of technological changes in the transformation of scientific research paradigms, root themselves in the soil of Chinese scientific research practice, seek innovation based on existing paradigms, and explore a scientific research path that is in line with China's national conditions and has distinct Chinese characteristics. Specifically, it is necessary to strengthen interdisciplinary cooperation and collaborative innovation, standardize data usage standards, promote the development of interpretable AI, and establish and improve research ethics review mechanisms, to provide solid theoretical support and action guidelines for innovative research paradigms.

Key words: scientific research paradigm; intelligent scientific research paradigm; artificial intelligence; data trap; crisis of confidence