AI参与科研，边界在哪里

『AI洗稿』『数据装饰』『代笔』等一系列新型科研诚信问题不断出现——

《光明日报》记者 崔兴毅

自大型语言模型问世以来，人工智能（AI）已广泛应用于知识获取、数据分析和学术写作等领域，显著提升了学习与研究效率。

但AI赋能科学研究的同时，也催生了“AI洗稿”“数据装饰”“代笔”等一系列新型科研诚信问题。以“算法参与”为特征的科研模式，导致了贡献界定模糊、研究过程不透明、结果可解释性下降等一系列新问题。更复杂的是，AI系统可能携带算法偏见，甚至生成虚假或误导性内容，这使研究结果的可靠性面临严峻挑战。

如何应对这些新挑战，已成为关乎科学事业健康发展的重要课题，记者就此进行了调研采访。

AI浪潮下，新旧问题渐次浮现

“AI对学术出版生态最核心的冲击，是科研生产力的大幅提升与滞后的科研生产关系之间矛盾的加剧。”清华大学图书馆馆长金兼斌观察到，一方面，从科学问题的提出与筛选到学术成果的产出，AI都在深度赋能科研；另一方面，从数据、图表造假到论文的AI生成，科研过程中的AI滥用日益严重，“学术诚信事件频发，论文撤稿大量增加。撤稿论文中有明显AI参与生成的比例显著高于非撤稿论文”。

在金兼斌看来，学术诚信并非新问题，只是在AI时代被进一步放大，有了新的造假途径和方式；但新问题同样存在——人们开展科研的方式正在被重塑。“比如文献综述环节，今后会越来越依赖AI完成。没有进入学科知识大模型语料的文献，对后续研究的影响将趋近于零。”金兼斌说。

上海交通大学医学院图书馆馆长仇晓春则从另一个角度感受到冲击：当AI参与论文的起草、构思甚至数据生成，传统的作者概念受到根本挑战，AI生成的文本模糊了抄袭与创新的界限；一些审稿人用AI撰写评审意见，导致评议质量下滑、内容趋同。“‘Publish or Perish’（不发表就出局）的循环加剧，此前侧重文献计量分析的学术评价体系面临风险。”仇晓春说。

对于这些问题，Frontiers（前沿）出版社首席执行总编范洺菘提出：“一个非常明显的趋势是，AI已经成为科学研究和新兴技术的基础设施。”在他看来，治理的起点在于区分“诚实使用AI”与“恶意使用AI”——很多非英语母语的科研人员用AI提升论文语言质量，“他们只是用工具改善了论文表达，这没有任何问题”；而用AI编造虚假内容冒充真实研究成果，则必须被识别和阻断。

还有一些新问题。“比如，AI的介入模糊了责任边界。大模型生成的错误或‘幻觉’内容，应被视为‘诚实的错误’还是归责于使用者？大模型能否作为作者或共同作者？”中国科学院院士梅宏在接受采访时举例说，“操纵AI审稿”等新伎俩已经出现，如在论文中隐藏“仅限正面评价”的指令，以欺骗AI评审系统。

算法的尺，能量得出真贡献吗？

技术与规则之外，更深层的是AI对科研评价体系的影响。

“‘唯论文’导向的学术考评制度，是导致各种学术诚信和学术出版问题的源泉。”金兼斌直言，基于发表量、引用量的量化考评，不断制造学者对论文发表的需求和大学对于排名的追逐，“学术出版过度商业化的倾向正在异化科学研究和学术发表的本质，掠夺性期刊、论文工厂、署名交易正是这种异化的直接表现”。

对此，金兼斌建议，大力加强本土优秀学术期刊建设，推进本土预印本平台和机构知识库建设；图书馆则应充分发挥掌握本校科研教学数据的优势，打破简单依赖发表量和引用量的考核模式，推动构建多元考评体系。

然而，更深层的挑战在于，“唯论文”评价体系已经难以适应AI for Science（人工智能驱动的科学研究）带来的科研活动多元化特征。中国科学院院士、北京大学教授鄂维南在接受采访时表示，科研成果的体现形式不应再局限于期刊论文这一“中间态”。

“一个革命性的想法、一套高质量的科学数据、一个被广泛使用的软件工具，其科学价值与社会影响力，可能远超一篇普通论文。”鄂维南透露，目前正在探索基于大数据的新一代评价方法，“利用‘玻尔平台’等新型基础设施汇聚的科研全过程数据，我们可以尝试对科研人员的‘想法—工具—数据—论文’这一全链条贡献进行更精细、更公正的度量”。

“这将有助于引导科研人员回归解决真问题的轨道，减少为发表文章而进行的重复性、跟风式研究，把宝贵的智力资源投入真正的原始创新和关键技术攻关上。”鄂维南说。

“唯论文数量、唯职称、唯学历、唯奖项，加上期刊影响因子，本质上都是考核人员用来回避深度评价工作的简便指标。”在范洺菘看来，科研评价本身就是非常艰难的工作，真正公平的评价，是深入理解一个人实际完成了什么工作，这项工作的真实价值是什么。

“而AI恰恰能在这里发挥作用，对不具备特定学科背景的考核人员而言，AI可以成为强大的对标工具，将一个人的工作与同领域的整体水平进行比较。”范洺菘同时提醒，任何工具都有两面性，如果训练数据本身无法覆盖科研全链条贡献，或者算法设计陷入另一种简化主义，那么看似客观的AI评分反而可能制造出更隐蔽的偏差，让评价失去应有的深度与温度。

数据的镜，能照见责任边界吗？

中国科学院院士谭铁牛警示：AI生成的虚假或错误内容，若未被有效识别而进入学术传播链条，长期危害不容小觑。“人工智能对于科学研究而言，既是赋能创新的‘阿拉丁神灯’，也可能是开启风险的‘潘多拉魔盒’。”谭铁牛说。

在此背景下，让原始数据可获取、可核验，便成了验证科研诚信的直接途径。

金兼斌认为，需要从制度与技术两个维度划定责任边界。他呼吁，应建立覆盖研究全周期的数据透明规范，要求科研人员在发表成果时，同步提交关键原始数据、算法代码与处理流程，并借助区块链等技术实现数据溯源和防篡改。

“当每一个数据点都可以回溯，每一次AI介入都有清晰记录，‘诚实错误’与‘恶意造假’的界限就不难厘清，责任就有了落脚点。”在金兼斌看来，越是AI深度参与的研究，越需要以透明赢得信任，以可复现性作为衡量研究可靠性的核心标尺。

数据核验是守住结果的底线，而学术出版的过程同样需要透明与协作。

范洺菘建议，应把评审重心放在论文本身，而非作者资历背景，“我们正探索将传统同行评审转变为审稿人与作者实时互动的协作式评审，为青年学者提供与资深科学家合作主持专刊、担任科学编辑的机会”。

这一过程中，AI可以扮演“公正助手”的角色。

“自动隐去作者身份信息，实时分析评审意见是否聚焦于内容而非资历，甚至辅助比对已有研究数据，让青年学者的创新不被盲目的偏见遮蔽。”范洺菘直言，“但工具终归是工具，协作式评审的真正温度，仍来自人与人的坦诚对话，来自前辈对后辈的托举与信任——这是算法无法计算的价值。”

审稿公平只是学术生态的一环，更根本的防线在于让评价体系回归人的判断。

“通过构建多元赛道、提供长周期支持、严守学风底线，努力将制度压力转化为学术动力。”仇晓春认为，图书馆也将迎来从“资源仓库”向“评价智库”转型的关键机遇——通过整合多维数据、提供多维度影响力报告，帮助评审专家从更多视角理解一项成果的价值。

“但报告提供的是信息而非判决，AI计算出的影响力分数永远无法替代同行对一项工作内在突破的深刻洞察。”仇晓春直言，“评价改革的核心，正是让AI成为照亮学术贡献全貌的镜子，而非裁剪人才与创造力的刀子——如此，技术才能真正服务于人，服务于科学那不可被量化的好奇心与远见。”

据2026年8月26日《光明日报》