AI当“科研搭子”，你真的放心吗？

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/42078.html

AI当“科研搭子”，你真的放心吗？
。近日，安徽财经大学通报了一起引发学界关注的学术不端事件：一篇论文注释中判决书序号被质疑编造，校方调查后认定，作者“误信”了AI的检索结果，属于AI使用不当行为，相关责任人受到相应处罚。

个案的处理有了结论，但它抛出的问题远未结束——当AI深度嵌入选题构思、数据分析，乃至论文成稿的科研全流程，效率跃升的另一面，是原创性归属模糊、黑箱责任难辨、代笔界限不清等深层困惑。

“AI科研搭子”成日常，一线科研人员如何在借力与守住学术底线之间找到平衡？记者采访了多位不同学科的研究者，听他们讲述真实的使用体验与边界思考。

借助AI输出科研思路，好点子算谁的？

“过去招研究助理时，我会比较看重编程能力和执行效率。现在AI把这部分能力的门槛大幅降低了。”中国农业大学经济管理学院教授朱晨说。今年春节前后，她搭建了自己的科研智能体系统，将经济学实证研究的流程拆解为7个环节，交由不同的AI代理分别执行。

变化是显著的。朱晨发现，智能体可以基于本地数据库自动生成候选研究问题。在她记录的14次运行中，系统共生成79个候选问题，87%符合变量存在、研究设计与数据结构匹配、计量方法可行这三个基本条件。“从形式上来说过关了。”她评价道。

但“过关”离“出色”还有多远？

一位长期关注中国宏观经济政策与微观基础领域的研究者，几乎浏览了苏黎世大学Yanagizawa-Drott团队公开的全部200余篇AI生成论文。他的评价相当克制：“只有极少数选题具备继续推进的价值，大多数只是刚入门的研究生水平。如果送到我手里，我会直接拒掉。”

问题在于分析深度。以经济学常用的双重差分方法为例，这位研究者指出，一篇成熟的实证论文不仅要完成基准回归，还需进行平行趋势检验、敏感性分析和异质性处理效应分析。但在目前公开的AI生成论文中，这些关键步骤“往往缺失，或者只是形式性地出现”。

这指向了一个更深层的问题：当选题本身都可以由算法生成时，科研的原创性到底从哪里来？

华大集团首席研究员徐讯将生命科学领域的AI能力分为五个级别。他判断，当前AI处于L3阶段——“ 它能完成发现，却不能自主提出原始问题，而提问仍是人类不可替代的核心价值。 ”在他看来，AI可以基于现有知识做组合与推理，但真正的科学问题，那些需要跳出已知框架、指向未知的追问，仍必须由人来完成。

西安交通大学网络空间安全学院教授栾浩认为，大模型本质上是一个知识搜索器。与传统关键词搜索不一样的地方在于，它能理解搜索词的意思，然后在已有的知识库里进行精准语义检索。这意味着，AI能做的是“ 精准全面地检索人类现有的知识库 ”，而“ 怎么去获取人类未掌握的新知识，大模型无法代劳 ”。在他看来，有时AI给出的思路看似很新，“ 其实并非创造，是有其他人提供过相关的思路，被它录入到自己的知识库了 ”。

清华大学化学系副教授马冬昕的实践感受印证了这一点。她把AI称为“ 一个拥有大量跨学科知识，会聆听、爱分享、随叫随到的好朋友 ”，但坚持认为课题选择不能完全交给AI。“ AI会给出一系列参考文献，旁征博引地给出分析。但限于现有技术水平，难免出现谬误。必须追溯原始文献，独立自主且具有批判性地思考。 ”马冬昕说。

她特别提到，和AI探讨科学问题本身就有价值——当AI“ 文不对题 ”时，恰恰促使她更清晰地提炼问题，“ 这个思考的过程，本身就可以帮助我理清思路 ”。

借助AI开展实验分析，若出错谁担责？

构想落地，进入实验与分析。这是科研最“ 硬核 ”的环节，也是AI介入最深入、风险最隐蔽的环节。

中国科学院自动化研究所研究员李林静观察到，科研智能体已经能够完成“ 文献研读、实验规划、参数迭代、结果分析的闭环自主科研流程 ”。在北京大学干细胞研究中心，一套占地6平方米的智能平台能独立完成从指尖采血到干细胞培育的120多个步骤，实验误差从最高10%降至1%以下。

效率令人振奋，但隐患同样真实。

谷歌旗下DeepMind曾用AI预测出220万种新的晶体结构，其中38万种被判定为“ 稳定 ”。然而，得到实验验证的不到0.2%。中国工程院院士李国杰将这一现象概括为：AI产生了海量“ 看似合理 ”的假说，绝大多数无法被验证，导致科研资源被浪费在低优先级选项上。

“ 看似合理 ”四个字，精准击中了许多一线研究者的焦虑。

中国科学院空天信息创新研究院国家对地观测科学数据中心常务副主任张连翀的观察很直接：“ 通用AI系统缺乏主动拒答的能力——即便面对超出其知识范围的问题，它仍会尽力生成一个看似合理的回答，而该回答的准确性并无保障。 ”

栾浩从技术原理上解释了这种“ 一本正经胡说 ”的根源：当问题超出大模型的知识库，或者大模

型对于问题的理解有偏差时，它会“强行搜索”，给出不相关甚至错误的答案。

南京航空航天大学副研究员伍群芳在电路设计中也反复遇到类似情形。他举例，AI生成的电路方案从框架上看“基本原理确实可以这样去工作”，但一旦涉及效率、纹波、动态响应等具体工程指标，“它不可能所有东西都兼顾得到”。学生拿AI给的滤波电路或信号调理电路去仿真，往往发现实际结果与纸面方案相去甚远，“给个思路可以，最后实际用的东西还是要改变很多，要自己去仿真、优化、调整”。

朱晨的应对之道，是强制复现，同时保留人在关键节点的判断权。在她的系统里，每次研究完成后，都会生成完整的R代码，研究者可以基于原始数据重新运行分析流程，确认回归结果与报告一致。“这一步复现是整个流程中不可或缺的环节，也是确保研究可靠、避免AI幻觉的关键。”

张连翀和团队有个类似的解决办法——探索一条“可信语料库+数据治理+检索增强生成”的解决路径。他们在后台搭建了经过严格审核的可信语料库，智能客服的所有回答均被限定在这一可控数据源内。

这种“人机协同但人最终负责”的模式，正在成为越来越多实验室的操作方式。但问题在于：当AI的输出是一个“黑箱”，研究者既无法完全理解算法内部逻辑，又被要求为结果承担全部责任，这个责任边界的划分合理吗？

“当下大模型在企业实际部署中推进缓慢，核心障碍正在于此。”在栾浩看来，大模型给出答案的准确性在理论上无法精准量化，是其在工业界大规模落地的根本瓶颈。科研场景与工业控制同理：AI可以辅助，但决策权必须留在人手里。

中国人民大学物理学院副研究员高泽峰划出的边界更具体。他鼓励学生用AI辅助写代码，但到了结果分析环节，要求学生独立完成。“必须对自己学术成果中涉及的所有数据负责，绝不能外包，自己当‘甩手掌柜’。”高泽峰说。

借助AI写论文，边界该划在哪里？

实验结束，进入成文阶段。这是AI介入最普遍、争议最集中的环节。某高校化学方向博士生向记者坦言，课题组使用AI的频率“近乎100%”，从选题、初稿到润色都已渗入。

“100%”的高频使用也放大了两个问题：代笔界限模糊，虚假参考文献泛滥。

“学生发来英文论文，句子语法规范、用词地道，却与学术表达不适配。”马冬昕遇到的情况颇为微妙，“我请AI译回中文，发现译文通畅。叫来学生一问，果然是先写好中文再翻译成英文。”

批评教育之余，马冬昕也在思考边界到底在哪里：“AI可以辅助搭建逻辑框架、梳理论文思路、修改语法、润色语言，但绝不能代笔。”而高泽峰则进一步要求学生，在论文撰写阶段同样必须独立自主完成，不能使用AI工具进行辅助。

作为学术期刊的审稿人，伍群芳从审稿人角度印证了这一点：“审稿的核心还是看创新点——相

比现有技术提高了多少？怎么提高的？技术方案是什么？最后验证的是什么？这些恰恰是AI无法替代专业判断的地方。”

栾浩用“画龙点睛”来概括人的不可替代性。“大模型可能可以帮你完成80%甚至90%的工作，但是画龙点睛的那一笔，往往做不到。”他用做饭来打比方：AI做出来的菜，像食堂里的大锅饭，按照经验和机械化流程完成，“这道菜毕竟没有办法喂到它嘴里，它不知道咸淡”；而人给家人做饭时，有观察、有感知、有反馈闭环，“他知道这道菜做出来是什么感觉”。在栾浩看来，科研中最关键的那一笔，恰恰来自这种只有人才能完成的“点睛之笔”。

依赖AI的代价不止于此。

中国环境科学研究院研究员阳平坚就发现，一些学生有了AI后自己不再认真读英文文献，“这对学生科研能力伤害非常大”。他现在要求学生文献必须自己读，读完后可以用AI翻译和总结。

“我一周至少看5篇跟我学术方向相关的论文，了解最前沿的指标。电力电子领域技术性特别强，效率能做到99%还是99.1%，这些我心里都有数。”在伍群芳看来，这种对领域核心指标的持续敏感，构成了判断AI输出可信度的基准线。

虚假引文是另一个更隐蔽的陷阱。

阳平坚举了一个例子：AI在一次内容梳理中输出“中国的碳市场即将启动”，课题组成员没有发现问题。“中国碳市场2021年就已经启动了。AI调取的资料库版本比较旧，没有收录最新的数据。”在他看来，如果研究者对自己领域的常识和前沿缺乏专业敏感，AI输出的错误就会悄无声息地进入正式文本。

前不久，学术预印本平台arXiv发布了一项对250万篇论文、1.11亿条参考文献的系统性审核，结果显示：仅2025年，arXiv、bioRxiv、SSRN和PubMed Central四大平台中就存在近15万条由AI编造的虚假参考文献。

专家分析，其根源在于生成式AI的运作方式——它倾向于生成“看起来像文献”的内容，而非检索真实存在的数据。

“很多专业领域的数据库不对通用式AI开放，DeepSeek或者豆包查不到最新的前沿文献。而且它生成的参考文献，有些实际上根本就不存在。”阳平坚忍不住叹息，“这类利用AI写成的文献综述能有什么意义呢？”

伍群芳以自身领域为例佐证了阳平坚的观点：“电力电子学科高度依赖IEEE、IET等专业数据库，但这些数据库‘有权限、要版权’，很多内容没有接入通用大模型。好的东西AI够不着，所以生成的内容也流于大众，没有真正创新可言。”

这也提示，解决AI虚假输出的关键或许不在模型本身，而在数据源的可控性——如果科研人员能够将经过审核的专业文献和自有数据“喂”给AI，而非任由它从开放网络随意抓取，虚假引文的风险将大为降低。

北京大学副校长初晓波的一段话，或许是对这场讨论最恰当的收束：“人类使用AI辅助论文写作，绝非让渡主体性，而是探索一种全新的科研分工——让AI去处理数据的广度，让人类来守住思

想的深度与价值的温度。”

（本报记者崔兴毅）

作者：崔兴毅 来源：光明日报