■ 时论与争鸣

DOI:10.15998/j.cnki.issn2097-6763.2026.05.013



AI 能够参与同行评议吗

郑伊贝,赵梓含
(南京师范大学　教育科学学院,南京　210097)

摘　要:作为“作者”的 AI 吸引了研究者的集中研判,然而作为“审者”的 AI 却越过了学术共同体的应然思考,以隐蔽方式存在。这是因为同行评议自身问题域的混乱与复杂性,系统地为 AI 的介入搭建好了表演舞台,同时“论文工厂”引致的“不对等困境”“质量危机”与“审稿人的过度想象”也为 AI 的参与铺垫好了所有剧情,只等待 AI 审稿人的到来。事实证明,目前尚不具备自我意志的 AI 只能作为评审的助手,无法替代人类审稿。为解释 AI 审稿人的有限性问题,需要以语言、思维与理解为着眼点,透视人类审稿人与 AI 审稿人的本质区别。一方面,人类审稿人具有历史性与意义的生成能力,需要具备综合“事实”与“价值”的分析能力;另一方面,无论是“作者”还是“审者”都需要具备对创造性知识的判别能力,而基于概率推断的 AI 审稿人尚不具备这些能力。AI 审稿人虽然是“神秘的客人”,但并非“恐怖的敌人”,以人类智能实质性关联人工智能,才能真正推动学术共同体的发展与思维的解放。

关键词:生成式人工智能;大语言模型;同行评议;AI 审稿人

[**中图分类号**]G644　[**文献标志码**]A　[**文章编号**]2097-6763(2026)05-0135-09

一、作为“作者”的 AI 与作为“审者”的 AI

自 2022 年年底以来,以大语言模型(Large Language Models, LLMs)为代表的生成式人工智能强势介入科学研究,在文本领域展现出强大的语言能力与科研潜质,正在开启有别于经验科学范式、理论科学范式等传统范式的“第五范式”[1]。人工智能(Artificial Intelligence,AI)正以科研辅助者的角色深度参与学术生产,但这种辅助并不必然确保产出“好”的学术研究成果,在显著提高科研产出数

修回日期:2026-01-21
基金项目:国家社会科学基金教育学一般项目“数字文明视域下教育空间生产的精神图景建构研究”(BAA250042)
作者简介:郑伊贝,女,四川眉山人,南京师范大学教育科学学院博士生,主要从事高等教育理论和教育哲学研究。
通信作者:赵梓含,男,江苏连云港人,南京师范大学教育科学学院博士生,主要从事高等教育理论和教育哲学研究。
引用格式:郑伊贝,赵梓含. AI 能够参与同行评议吗[J]. 重庆高教研究,2026,14(5):135-143.
Citation format: Zheng Yibei, Zhao Zihan. Can AI participate in peer review[J]. Chongqing Higher Education Research,2026,14(5):135-143.

量的同时,还可能导致平庸论文的激增[2]或学术泡沫的产生[3],甚至成为人文社科学者们的替代者[4]。本文无意强调人工智能对不同学科领域的赋能或风险,而是以 AI 作为作者介入科研领域所引发的学界关注与审慎接纳为切入点,进一步展开对 AI 参与同行评议这一问题的探讨。

在 AI 广泛且深度介入科研的同时,学界展开了大量相关研究。相较于 AI 带来的科研机遇,研究者更加关注如何防范 AI 带来的风险以及如何对研究者使用 AI 进行规制[5-8]。就研究主题来看,既有文献还存在以下遗漏或偏见:一是对研究者群体选取的偏狭。相较于成熟学者,科研新手特别是研究生群体更容易成为学界关注与规制的对象。对研究生群体滥用、误用 AI 的防范规制是为了帮助他们成为更好的研究者,或避免更深层的科研灌水,但也可能忽略其他更值得关注的议题,如成熟学者群体在进行科研时的 AI 使用情况、使用方式以及使用经验等。事实上,无论是分析不同研究者群体在 AI 参与科研上的差异,还是将 AI 置于整体的学术生态中考量,都需要对成熟学者予以更多关注。否则,我们有理由怀疑这是学界对研究生群体的一种集体性怀疑与偏见。二是已有研究格外关注作为作者的 AI。当前,各种 AI 多处于通用大语言模型阶段,尚未在人类社会各领域实现精细化或实质性深度参与,其介入多集中于文本生成、信息处理等环节。由于学术研究仍受成果导向与学术锦标赛[9]等评价制度的影响,论文撰写与发表逐渐成为科研的重要呈现方式,这能在一定程度上解释为什么学界对 AI 介入科研议题的思考更加关注作为作者的 AI。因此,无论是斯坦福大学发起的 Agents4Science 2025“首个要求 AI 同时作为论文主要作者与评审者的开放会议”[10],还是华东师范大学组织的“AI 驱动教育研究论文写作”[11]征文活动,无一不在呈现学界对作为作者的 AI 的关注。

以上研究旨趣的偏见,致使我们遗漏了对 AI 介入科研领域这一复杂性问题的研判。在作者之外,AI 还会以何种身份以及如何介入科研? 在 AI 引发的作者问题受到集中热议的同时,AI 衍生的审者问题已悄然而至。2025 年 6 月,韩国学者洪成镇通过对审稿意见中语言模糊、逻辑矛盾及缺乏同理心等特征的分析,在《泰晤士高等教育》撰文怀疑其“论文极有可能被 AI 评审过”[12]。若个案尚不足以说明问题,2025 年 11 月的国际学习表征会议(ICLR)则提供了大规模的客观实证,超过 21% 的评审内容完全由大模型一键生成,35% 的意见则经过 AI 的不同程度润色,真正由人类撰写的比例仅剩 43%[13]。总而言之,除了作为“作者”的 AI,AI 已经以“审者”的身份深度介入学术生产环节。不过,作为审者的 AI 是如此隐蔽以至于尚未得到研究者的高度关注,一个正在参与论文评审并掌握裁决发表权的角色却处于无身份且不担责的境况之中。借用尼采(Friedrich Nietzsche)将虚无主义比作“所有客人中最阴森可怕的客人”[14]的隐喻,作为审者的 AI 同样是学术界最为隐蔽且影响重大的“客人”。一方面,作为审者的 AI 虽已事实性在场,却尚未获得学界的充分关注与审视,成为一个不被承认的角色,而这个角色可能引发评审者的身份模糊以及干扰科研质量评判秩序等一系列伦理问题;另一方面,即使作为审者的 AI 已较大范围参与到同行评议中,却因其隐匿性引发一种例外状态,即相较于对 AI 作者严防死守的规制,学界对于同行评议中的 AI 审稿人却往往选择性忽视。这意味着,AI 参与同行评议问题已经越过或逃逸出学术共同体的应然思虑。

二、AI 参与同行评议的系统性前奏

事实证明,学界对科研领域作者或科研前端的关注度远超审者或科研中端。本文基于此进行猜想,同行评议自身可能存在某些因素,共同构成 AI 以审者身份介入科研领域的系统性前奏。对这一问题的探讨不能匆忙地对既有事实展开批判性补足,而是要先分析作为审者的 AI 何以参与同行评议。因此,我们需要回顾同行评议的原初目的与现实境遇。

(一)同行评议自身问题域的混乱与复杂

发端于欧美地区的同行评议历经数百年之久,已融入世界各国职业学术圈的职称评定、论文发表

与课题资助等环节。依照其原初理念,科研成果需通过资深或至少是成熟同行的审视与评判。这一过程不仅能够促进知识和思想生产,还能够通过评议式学术对话凝聚共同体的科研共识、提升学术精神。尽管有学者基于个体的自主探究权力、学术共同体内部集体权力以及来自外部行政权力的博弈分析,认为同行评议并非最优的制度设计,只是达成 3 种权力博弈均衡的最不坏的途径[15],但我们依然需要承认其所承载的学术追求。

理念中的同行评议在实践及各国制度移植中暴露出诸多弊端。从实践来看,同行难觅[16]、审稿意见造假[17]、科学伪装[18]、评审代笔[19]等乱象引发了学界对其公信力的质疑乃至排斥。就制度移植而言,同行评议在引入我国后因多种原因产生某种异化,“权力支配、人情主导、标准缺失,只要居其一,都会使学术评价的意义受到严重损害”[20]。出于对构筑更好学术伦理世界的期望,现实中同行评议在不断受到批判与竭力改进中努力实现其理想中的存在价值。“这个制度之所以能得以长期坚持,在很大程度上是因为国际学界不断对其操作规程和相关配套制度加以完善。”[21]

可以发现,同行评议本身就是一个不断衍生各种问题、混乱且复杂的问题域。新兴技术 AI 凭借对人类自然语言的拆解与重组成为一种催化剂,其介入更是加剧了同行评议问题域的复杂性与模糊性,客观上也为其自身以审者身份隐蔽入场提供了现实空间。

(二)论文工厂与对理想审稿人的过度想象

如果说上述只是同行评议自身的问题,那么更重要的是,全球范围内科研的“工程化”“成果化”“功利化”等价值导向,连同高等教育系统的大规模扩张,将科学研究异化为迫切需要 AI 审稿人的论文工厂。

论文的发表诉求带来了作者与审者规模的不对等。相较于科研人员数量的指数级上升,合格审稿人的扩容明显滞后。审稿任务的过载引发审稿人的审稿疲劳,“随着投稿量的持续增加,被召集审稿的学者也开始负担过重。此外,提交的论文质量低且重复性高,导致过程单调,审稿人感到疲惫”[22]。这种不对等困境进一步诱发同行评议的质量危机。建立在学者的学术责任与学术道德之上的同行评议,在时间成本与精力投入的双重挤压下,不可避免地会面临评审意见的质量滑坡。考虑到审稿人自身的学科背景、知识储备与各种非理性因素对论文评审的综合性影响,同行评议完全有可能演变为科研“公地悲剧”。迟滞的评审周期与加速的知识生产之间形成巨大张力,催生学界对理想审稿人的过度想象。这实质上是对不知疲倦且高效评审的审稿人的幻想,“时间不足是使用 LLMs 撰写评审意见的原因之一,如今,面对海量邀请,审稿人早已应接不暇”[23]。事实上,对于作者与审者数量的不对等以及评审质量的滑坡问题,我们还能希冀学者的责任感与学术操守予以对冲,而当科研遵循加速社会中的资本增值逻辑,人类智能已难以适配论文工厂不断提速的学术生产,便只有高效、稳定且快速的人工智能才能担此重任。

综上所述,同行评议系统内在的问题域与复杂性为 AI 的介入搭建好了表演舞台,“不对等困境”“质量危机”与“审稿人的过度想象”为 AI 的介入铺垫好了所有剧情,只等 AI 审稿人的到来。审者在传统科研中本就具有隐蔽性,AI 参与同行评议后继承并强化了这种隐蔽性。相较于对作者的严格规制,学界对 AI 审稿人不愿正视与承认的态度,实质上在无意识中促成了学术责任的部分让渡。这也是近年来开放科学特别是开放同行评议出现的主要推动力。鉴于这并不作为本文的核心议题,在此不过度展开。

三、AI 审稿人的角色期待与评审表现

文献梳理发现,关于 AI 审稿人的探讨尚未引起学术界的充分关注。2025 年 12 月,国内多家学术期刊编辑部陆续发布规范使用生成式人工智能技术的有关说明或规定,均主张以积极审慎的态度

回应 AI 介入科研领域。其中,针对作者问题普遍强调公开、透明、诚信,针对审者问题多采取严厉禁止或有限使用的立场,如“均不得在审稿与编辑环节使用生成式人工智能工具撰写审稿意见或处理稿件内容,严禁将投稿论文上传至任何公开的生成式人工智能平台”[24],“仅允许有限使用本地部署的 GenAI 工具进行辅助审稿”[25]。换言之,所谓的积极审慎实质上凸显一种差异化对待,即积极、宽容对待作者问题,而审慎、严肃对待审者问题。

(一)人工智能自动化评审与智能化评审

如前所述,同行评议的内在困境构成 AI 审稿人出场的系统性前奏。由于效率低下和审稿人数量有限,导致作者获得的反馈有限,阻碍作者及时获取审稿意见[26]。事实上,早在各类 LLMs 广泛应用于科研之前,2018 年国际上就有学者提出“人工智能同行评议时代”[27]的构想——自动化软件帮助审阅论文,最终决策权仍掌握在人手中。由于当时的算法尚不成熟,只能辅助格式审查、核对参考文献等前同行评议工作,因此只能算是自动化评审而非智能化评审。

2024 年,一篇题为《生成式人工智能时代的同行评议》的研究进一步指出,相较于以往的自动化标准审查,GPT-4 在撰写和整合审稿意见方面具有中等潜力。“对于评审流程中需要较少智力投入或专业领域知识的常规环节,如可复现性检查、可读性评估、改进评审报告的撰写等,AI 的作用尤为显著。”[28]总体而言,目前的 AI 审稿人虽然存在幻觉、创新性识别不足等问题,但我们无法选择性忽视 AI 参与同行评议的未来潜力。

随着 LLMs 的能力跃升,AI 审稿人已从理论构想走向现实实验。有研究者试图构建一个多角色、多智能体的评审框架,旨在生成全面、准确、逻辑连贯且符合人类评审判断的评审意见。他们“构建了一个新颖的数据集 Review-CoT,其中包含 14.2 万条评审意见,专为训练大语言模型智能体设计。通过一种相关论文感知的方法,训练出能进行结构化推理的大语言模型评审智能体 Review Agents”[26]。凭借对文本信息的高效分析、处理、概括以及对人类自然评审语言的标准化生成,LLMs 能够以相关性概率推断能力模仿人类评审。更重要的是,相较于一味禁止,我们亟须客观、全面审视 AI 审稿人的各项能力表现。

(二)作为辅助的 AI 审稿人与学术共同体的伦理扩容

AI 审稿人展现了何种能力?相较于人类审稿人又欠缺什么?AI 在同行评议及其相关环节中展现出不同类型的能力。一方面,AI 审稿人更能满足科研效率的诉求,进而缓解人类审稿人的审稿疲劳。研究表明,“AI 可以在分钟级输出评审文本,而人类往往需要数周”[29],“AI 能够辅助结构化总结、语言与表述改进等,在一定程度上减轻评审负担”[30]。另一方面,AI 审稿人的评审意见需要人类审稿人最终把关。对于一些专业知识领域,“AI 在速度、一致性、覆盖面与结构化总结方面具备优势,人类审稿人则在语境理解、创新识别、价值权衡与责任承担方面不可替代”[29]。人类审稿人如何处理 AI 审稿人给出的评审意见,将在很大程度上影响评审质量。此外,AI 参与审稿还涉及知识产权、学术隐私等伦理问题。有研究对不同学科领域且在性别、年龄与学术职称上具有多样性的多位审稿人进行问卷调查与访谈,受访者指出 LLMs 可能存在“潜在的偏见、缺乏透明度、隐私和保密风险”[31]。对此,有研究者基于对 Facebook 中“2 号审稿人”大型学术社群关于 AI 审稿人的评论分析,提出了在同行评议中使用 AI 的建议,包括“增加透明度、加强质量保障、确保伦理规范、保护隐私以及支持审稿人的持续专业发展”[32]。

在实际操作中,AI 审稿人目前能胜任的是“职业化编辑”而非“学者型编辑”的主要工作,具体包括参考文献、图表细节等论文格式与规范性检查、字词误用或语法问题等可读性评价以及论文核心内容的概括与关键信息的提取。就此而言,AI 审稿人完全能参与一些结构化的表层评审工作。对于人类评审专家与“学者型编辑”而言,AI 审稿人虽然能以海量的知识库支撑意见生成,但目前能实现的

是一致性的意见汇总,而非复杂性内容的理解与批判。当然,AI 审稿人还有其他能力,但大都能够归属为高效生成的能力。

AI 审稿人同样存在许多局限。科研泄密、知识产权等并非核心问题,因为传统评审也存在类似弊病;模型幻觉等评审意见编造现象不是关键问题,或可通过技术调优予以解决;同质化评审意见也不是重要问题,因为缺乏多样性与建设性意见严格意义上归属于算法局限,通过本地部署与个性化调试同样可解。AI 审稿人的主要问题在于创新性识别能力的缺陷。受限于知识库的更新滞后以及深度科学推理的缺失,AI 对真正新颖的研究往往难以充分识别,只能生成大而全的空泛意见。人类专家对思想创造力的判定与问题批判力的识别才是评审的核心所在。对于学术研究而言,正确且平庸的研究的价值在理论上绝不会超过有瑕疵但有趣的研究。换言之,创新性识别方面的无能,决定了所有其他能力均只能作为辅助。总之,目前尚不具备自我意志的 AI 无法替代人类审稿人,也无法作为评审代理人,只能作为同行评议中的补充与辅助。

综上所述,评审的效率、全面性与基本内容的理解能力是 AI 审稿人的优势,本质上 AI 审稿仍是人工智能依赖大数据、大算力、大算法对论文进行数据化拆解与概率性呈现。受限于自我意识、具身经验与生活世界认知的缺乏,当前的 AI 难以企及人类深度的语义理解、创新层面的判别以及价值层面的选择。当务之急是将 AI 审稿人从科研领域的灰色地带中辨识出来,正视其参与同行评议这一尚未得到足够关注的既成事实。唯有如此,我们才能推动同行评议制度的自我革新,并在真正意义上促进高质量的学术产出。总之,无论人类承认 AI 审稿人与否,学术共同体已然在事实上对其进行了伦理层面的扩容与接纳。与其说我们期待一个更完美的 AI 审稿人,不如说是在呼唤一种更健康纯粹的学术生态。

四、AI 审稿人的语言、思维与理解

既然目前的 AI 审稿人无法进行实质性的评审,仍需人类审稿人最终把关,因此我们需要进一步反思 AI 审稿人何以“可能”与“不能”。AI 审稿人对论文的拆解与重组并不等于真正理解,基于数字或算法语言的评审意见,在本质上区别于人类自然语言的评审意见。LLMs 的底层逻辑是以经验性的概率推断来模拟人类的思维方式,故而 AI 审稿人缺乏人类审稿人的深度理解与因果思维能力。因此,对 AI 审稿人的审视需要以语言、思维与理解为着眼点。

(一)人类文字语言的历史性与同行评议的意义生成

人类文字本质上是兼具可见性与历史性的表达、沟通与理解的符号系统。一方面,文字的可见性使其能超越时空的“当下性”限制,成为人们真实可感的认知与信息工具。表音文字继承源自声音语言的特质,实现可视即可读,而表意文字(如象形文字)则以真实世界或想象画面作为文字构造基础,实现可视即能想。另一方面,文字以一种文明形式或历史样态进行代际传递,并在形态流变或隐喻象征中逐步形成民族或地域特色。这种历史性强调语言对话的意义生成,而意义联结着人与人、人与物,不仅是人类生活必不可少的文化产物,更是构筑精神世界的必要根基。

人类审稿同样如此。评审意见并非对研究成果纯粹客观的审视,而是不可避免地融入审者的主观性,这种主观性正是作者与审者之间意义生成的必要条件。从还原论的视角来看,无论是声音还是文字形式的语言均是人们对内在意识的表达。内在意识是主体在意识到“内在的自我”与“外在的世界”后产生的一种内外分别,既是“思”的形式,也是对“思”的进一步“反思”。因此,无论是“思”或是“反思”,人们都需要借助概念这一思维的普遍性对象,通过感性认识达致知性表达,最终进行理性思维运作是人认识世界的普遍方式。这种认识能使人与世界之间的关联获得意义表达。概言之,文字语言作为重要媒介,最终帮助人们建构起意义世界。

AI 审稿人借助数字语言呈现文字语言,本质上是通过暴力计算实现对论文核心内容的提炼与主要观点的拆解。人类与 AI 对语言的使用是不同维度的能力,AI 的数字语言及其背后的数理逻辑是一种普遍必然性的关系呈现。虽然文字与数字均归属于语言范畴,但前者致力于构建意义世界,后者推导甚至建构一个概率世界。AI 审稿人本质上涉及人类的语言问题,基于数字语言的 AI 审稿人对生成意义的无力,直接决定了它的有限性。

(二)人工数字语言的概率性与同行评议的最终判定

数字化时代的隐喻是数字语言已成为认识、理解与建构世界的基底。不同于传统社会的文字语言,数字语言正在不断挤压文字语言的出场空间,人、物与事皆可通过数字化转换成为基于数学运算与算法测度的存在物。实际上,AI 审稿暗含着一种概率性,即对事物出现可能性的推演与计算,因为数字语言更加侧重事实性与可能性判断。因此,AI 审稿人的评审意见是一种基于概率推断与频次复现的最大可能文本,其通过对语料库中评审意见的字词进行排序模仿与对论文的暴力拆解生成统计学意义上的关联,追求的是与既有研究的相似性而非超越既有研究的创新性。当我们评判研究具备价值时,作者与审者进行了主体间的思维碰撞与精神交流,因而具有意义性。字词是人们进行语言沟通与传递思想的基本单元,是将某事、物命名或概念化的载体。AI 审稿人基于概率性的评审容易导致其语言生成的确定性,而这种确定性将会引发“平庸的泛滥”与“创新的扼杀”。真正的思想或有趣的研究往往依靠不确定性或艺术性的概念创制获得“标识性”与“原创性”,只有不断创制此类概念才能构建自主知识体系。

因此,AI 审稿人本质上还涉及休谟(David Hume)难题,AI 即便能全面提取并生成论文的相关信息,依然无法进行最终裁定。如一篇论文虽有瑕疵,但其新颖视角仍可能成为人类审稿人眼中值得录用的理由。如果将人类审稿的标准以算法的形式注入 AI,赋予 AI 审稿人以人类的评判尺度,构建这样的 AI 评审专家系统的确存在技术可能,但这本质上仍是人类审稿与 AI 审稿的复合体。概言之,人类审稿需要在事实与价值间进行综合分析,论文的语言可被转化为数据化的事实呈现,而研究是否具有价值以及具有何种价值是 AI 无法判定的。在价值抉择面前,AI 审稿会陷入“算无可算”的尴尬境地,这也解释了为何学界始终强调人类审稿人的最终判定。

(三)思维的自反性与 AI 审稿人的关键性缺陷

AI 的语言呈现能力是对 token 的组装与下一个 token 的预测,而人的语言表达能力基于思维的自反性得以实现。这种自反性是指“思维要知道自己到底想了什么,就必须反思自身,将思想自身对象化于语言中”[33]。这意味着,文本信息从人类自然语言向人工数字语言的转变过程中存在着“思”与“反思”的缺失,这种缺失会造成思想性的“缩水”,即思维、语言、算法三者之间传递存在着“失真”。科研是知识的创造与思维的想象,无论是作者还是审者,都必须具备对创造性知识的判别能力。“人类的创造性是一种复杂的涌现,包含着超越大数据概率论和统计学原理的某种反常建构,本质上是反概率的。”[34]这在一定程度上佐证了 AI 审稿人的局限。不过,这种局限是人类眼中的局限,AI 审稿人的确能生成人类眼中的有效语言,赵汀阳形象地将其描述为“AI 的‘思维’仅限于处理数据,在海量标识间按照最大可能的关联性建立最优组合输出,人类之所以读到知识和思想,仅仅是因为人类拥有‘密码本’”[35]。因此,与其追问 AI 审稿人究竟还存在何种关键性缺陷,不如反思当前的学术研究存在何种缺陷。AI 审稿人在创新与反思上的缺陷,恰恰是对做“真学术”的警醒。

五、同行评议的智能关联与 AI 审稿人的身份重估

事实上,人类的理解是一种理性主义与经验主义的调和存在,在一定意义上也包含概率推测。当前 AI 的发展正依循这种模式,如果不能突破技术瓶颈获得“自我意识”,实现从“专用”到“通用”的

转变,那么就只是通过数字语言模拟和重组人类的经验世界,相当于“看见”但无法“看懂”。即使 AI 具备“自我意识”,但受限于硅基的存在方式,也未必具备像人一样的“先天综合认识形式”——感性接受能力、知性范畴分析机制与理性判断的综合框架①。换言之,我们无法在真正意义上要求 AI 像人,只要 AI 能够理解并回答人类的问题、与人类对话,便满足了原初的“图灵式”判断条件。同样,如果 AI 审稿人能够节省人类的评审时间、提高评审效率、缓解评审疲劳,那么就已经极具吸引力了。

AI 的潜能究竟能达到何种程度?如康德(Immanuel Kant)所承认的,人类只能认识“现象界”而非“物自体”,未来的通用 AI 或超级 AI 恐怕也不能被人类真正理解或认识。同样地,我们需要在承认人类理解限度的前提下,重新审视 AI 的经验世界与语言呈现机制。在此意义上,无论是当前还是未来,AI 审稿人可能依然会保持其神秘身份。概括来说,从技术或工具角度思考 AI 审稿人能够帮助我们廓清其能为与不能为的事实性边界,而从语言与理解的视角思考,AI 审稿人则帮助我们透视数字语言如何评审论文、分析人类审稿人与 AI 审稿人的本质区别,最终解释 AI 审稿人的局限性所在。

AI 参与科研实质上是一个跨学科或去学科的问题域,是以 AI 为圆心、以所有学术研究为半径展开的全覆盖式问题创生。LLMs 给科研领域带来的危机首先是“语言之死”,“语言之死”则可能进一步带来“思想之死”。各种意义上的“死”并不可怕,可怕的是对“死因”的未知。

首先,AI 审稿人的神秘性在于其隐蔽性,即一个正在参与论文评审且具有裁决发表权的角色却处于“无身份”且“不担责”的境况之中,其既可能损害同行评议乃至科研领域的基本秩序,也可能为同行评议制度的革新提供契机,这取决于学界对 AI 参与科研持何种认知立场。这也是有必要探讨“学术共同体的伦理扩容”的关键所在——如果是以促进人类文明的知识创造为立场,那么我们没有任何理由排斥作为审者的 AI。其次,当前的 AI 审稿人是尚不具备自我意志的“类审稿人”,其能力定位于同行评议的辅助,评审意见也只能“客(AI 审稿)随主(人类审稿)便”。再次,AI 审稿人作为“客人”具有不可忽视性,学界不能选择性无视。为此,我们需要思考,重估同行评议的价值是为了论文发表还是为了知识创造与思维绽放?重估 AI 的价值是为了让发文变得更加简单还是为了推动学术共同体的实质性发展?

总之,AI 审稿人虽然是“神秘的客人”,但并非“恐怖的敌人”。本文之所以提出“恐怖的敌人”这一参照,意指学界在认识到 AI 审者之后,不宜简单沿用 AI 作者问题的处理逻辑,因为 AI 作者的署名资格、原创性等问题尚未获得真正许可。与 AI 作者问题不同,AI 审者并不直接主张作者身份或原创资格,因此在一定程度上避开了这种许可权问题。更重要的是,从原创性与标识性等角度来看,只有真正的思想与知识创造才能成为未来人类科研的主要内容,AI 的自动生成往往会导致平庸的泛滥而非思想的创新。未来文明只有在真正意义上以人类智能关联人工智能,才有可能获得最大程度的思维解放。

参考文献:

[1] 颜世健,喻国明. 智能方法作为“第五范式”:人工智能时代科研范式的“新物种”[J]. 学术探索,2024(1):34-43.

[2] 张梦然. AI 提升科研产出但平庸论文激增[N]. 科技日报,2025-12-29(04).

[3] 杨亚红,余党会. 人工智能伴随的学术泡沫及预防策略分析[J]. 编辑学报,2025,37(1):27-31.

① 此处的“先天综合认识形式”源自康德在《纯粹理性批判》中确立的先验唯心论框架。康德认为,“无感性则不会有对象给予我们,无知性则没有对象被思维”。人类真实的“理解”必须依赖一套先天的认知结构:首先通过感性的时空直观形式无条件地接受外部质料;其次通过知性范畴对杂多的表象进行综合;最后通过理性进行理念层面的调节与统摄。AI 大语言模型本质上是对既有语言符号(token)的概率重组,缺乏植根于人类主体性、实现“从感性到知性再到理性”自发综合的先天认识机制。

[4] 人工智能闯入人文社科:学者们的“新助手”,还是“替代者”?[EB/OL].(20250228)[20251231].https://www.cssn.cn/skyl/skyl_skrp/202503/t20250306_5856124.shtml.

[5] 蔡芬,贾枭,沈文钦.生成式人工智能在我国研究生学术写作中的应用现状及其影响[J].中国高教研究,2025(1):75-82.

[6] 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.

[7] 印波,庄新宇.生成式 AI 在学术写作中的运用:科研诚信规制与伦理进路探索[J].自然辩证法研究,2025,41(7):100-109.

[8] 和鸿鹏.如何规范生成式人工智能的学术使用——基于 55 所世界一流高校的调查[J].中国科技论坛,2025(8):129-135.

[9] 阎光才.学术等级系统与锦标赛制[J].北京大学教育评论,2012,10(3):8-23,187.

[10] Open Conference of AI Agents for Science 2025[EB/OL].[20251231].https://agents4science.stanford.edu/.

[11] 第一作者必须是 AI!华东师大首发 AI“主笔”论文征集[EB/OL].(20251007)[20251231].https://mp.weixin.qq.com/s/Jb1QN639SV7DgBmd1XSQJw.

[12] Hong S. My paper was probably reviewed by AI – and that's a serious problem[EB/OL].(20250624)[20251231].https://www.timeshighereducation.com/opinion/my-paper-was-probably-reviewed-ai-and-thats-serious-problem.

[13] ICLR2026:AI 大模型审稿危机引发学术界反思与新规制定[EB/OL].(20251128)[20251231].https://www.sohu.com/a/959460253_121924584.

[14] 弗里德里希·尼采.权力意志[M].孙周兴,译.上海:上海人民出版社,2018:168.

[15] 阎光才.学术共同体内外的权力博弈与同行评议制度[J].北京大学教育评论,2009,7(1):124-138,191-192.

[16] 李澄锋,陈洪捷.知识生产方式的转型与同行评议的危机[J].高等教育研究,2020,41(12):22-29.

[17] 万志超,蔡静雯,姜海,等.国际同行评议中审稿意见造假现象及相关的学术不端防范[J].中国科技期刊研究,2021,32(5):571-575.

[18] 查自力,樊秀娣,朱晨光.为何学术同行评议反而不专业?——“科学伪装”的制度逻辑[J].江苏高教,2023(1):7-16.

[19] 王景周,范晴.学术期刊同行评议代笔行为的表现、原因及对策[J].科技与出版,2024(9):19-25.

[20] 李剑鸣.自律的学术共同体与合理的学术评价[J].清华大学学报(哲学社会科学版),2014,29(4):73-78.

[21] 何云峰.学术发表中同行评议的伦理基础[J].河南大学学报(社会科学版),2020,60(3):142-148.

[22] Swati P. Reviewer fatigue is real[J]. Indian Journal of Ophthalmology,2024,72(Suppl 5):S719-S720.

[23] Zou J. ChatGPT is transforming peer review-how can we use it responsibly? [J]. Nature,2024,635(8037):10.

[24]《中国人民大学学报》关于规范使用生成式人工智能工具的声明[EB/OL].(20251211)[20260102].http://xuebao.ruc.edu.cn/CN/news/news522.shtml.

[25]《科研管理》编辑部关于在论文写作和评审过程中规范使用生成式人工智能技术的有关规定[EB/OL].(20251231)[20260102].https://www.kygl.net.cn/CN/news/news10682.shtml.

[26] Gao X, Ruan J, Zhang Z, et al. ReviewAgents: bridging the gap between human and AI-generated paper reviews[EB/OL].(20250311)[20260102].https://arxiv.org/abs/2503.08506.

[27] Heaven D. The age of AI peer reviews[J]. Nature,2018,563(7733):609-610.

[28] Atreyi K. Peer review in the age of generative AI[J]. Journal of the Association for Information Systems,2024,25(1):76-84.

[29] Farber S. Comparing human and AI expertise in the academic peer review process: towards a hybrid approach[J]. Higher Education Research & Development,2025,44(4):871-885.

[30] Mohamed A A,Daniel C,Yang J, et al. Leveraging artificial intelligence in the peer review of neurosurgical research articles[J]. Neurosurgical Review,2025,48(1):631.

[31] Saman E, Nejadghanbar H, Salman A R, et al. Exploring the impact of generative AI on peer review: insights from journal reviewers[J]. Journal of Academic Ethics, 2025,23(3):1383-1397.

[32] Louie G. Benefits and challenges of using AI for peer review: a study on researchers' perceptions[J]. The Serials Librarian,2024,85(5/6):144154.
[33] 赵汀阳.语言开启思维:语言奇点与思维初始基因[J].当代语言学,2024,26(6):825841.
[34] 赵汀阳.意识的自反性:知者与作者之惑[J].中国社会科学,2025(8):6181,205.
[35] 赵汀阳.人工智能还给人类的思维难题[J].中国社会科学,2024(8):104123,206.

(责任编辑:杨慷慨　齐春梅　校对:齐春梅)

Can AI Participate in Peer Review

Zheng Yibei, Zhao Zihan
(*School of Educational Science*, *Nanjing Normal University*, *Nanjing* 210097, *China*)

Abstract: While AI as an "author" has attracted considerable scrutiny from researchers, AI as a "reviewer" has bypassed the normative considerations of the academic community and remains concealed. This is because the intrinsic confusion and complexity within the problem domain of peer review have systematically set the stage for AI's involvement. Simultaneously, issues arising from "paper mills"—such as the "asymmetry dilemma", the "quality crisis", and "reviewers' excessive imagination"—have created conditions for AI's participation, awaiting the emergence of AI reviewers. Evidence indicates that current AI, lacking autonomous will, can only function as an assistant in peer review and cannot replace human reviewers. To elaborate on the limitations of AI reviewers, this paper examines the essential distinctions between human and AI reviewers from the perspectives of language, cognition and comprehension. On one hand, human reviewers possess a generative capacity for historicity and meaning, requiring an analytical ability to synthesize "facts" and "values". On the other hand, both "authors" and "reviewers" must possess the capacity to discern creative knowledge, which probability-based AI reviewers currently lack. Nevertheless, while AI reviewers may be "mysterious guests", they are not "terrifying adversaries". Substantively linking human intelligence with artificial intelligence is essential for promoting the development of the academic community and intellectual emancipation.

Key words: generative artificial intelligence; large language models; peer review; AI reviewers