doi. 10. 3724/1005-0566. 20261007

技术—社会共构视角下 GAI 驱动知识生产的风险构型与协同治理研究

丁大尉[1]，李正风[2]

（1. 烟台大学文学与新闻传播学院，山东　烟台　264005；
2. 清华大学科学技术与社会研究所，北京　100084）

摘　要：生成式人工智能（GAI）正深刻重构知识生产的技术逻辑与社会结构。“主权 AI”时代“技术—制度—责任”三元结构的内在张力以及 AI 知识生成中的潜在风险与治理，已成为事关 GAI 演化与发展的重要问题。本文以语言互动类大模型为例，在分析其知识生产本质的基础上指出，生成式知识已呈现出明显的“拟主体化”特征，并且伴随着知识异化、知识“幻觉”、知识权力化、知识语境集中化等风险构型。在 AI 数据来源、算法设计以及自主评价和决策均内嵌人类价值的情况下，必须建立正确的人机关系以有效规避可能的知识生成风险。研究认为，未来知识生产中的人机共建须保证人类的主体性和创新性地位，同时必须建立异质性行动者共同参与的融合认知正义的综合治理框架，以保证 GAI 技术逻辑符合人类社会发展的内在需求。

关键词：GAI；知识生产；AIGC；人机协同

中图分类号：G301　　**文献标识码**：A　　**文章编号**：1005－0566（2026）10－0080－07

Research on the risk configuration and collaborative governance of knowledge production driven by GAI from the perspective of technology-society co-construction

DING Dawei[1], LI Zhengfeng[2]

（1. *School of Literature and Journalism*, *Yantai University*, *Yantai* 264005, *China*;
2. *Center for Science*, *Technology and Society*, *Tsinghua University*, *Beijing* 100084, *China*）

Abstract: Generative Artificial Intelligence (GAI) is profoundly reshaping the technical logic and social structure of knowledge production. The inherent tension within the "technology-institution-responsibility" triad structure in the "sovereign AI" era, as well as the potential risks and governance in AI knowledge generation, have become important issues concerning the evolution and development of GAI. Based on the analysis of their knowledge production essence this article takes the large interaction-based language models as an example and points out that generative knowledge has shown obvious "pseudo-subjectivity" characteristics and it is accompanied by risk configurations such as knowledge alienation, knowledge "hallucination", knowledge powerification, and knowledge context concentration. Under the circumstances where AI data sources, algorithm design, autonomous evaluation and decision-making are all embedded with human values, it is necessary to establish a correct human-machine relationship to effectively avoid possible knowledge generation risks. The research suggests that in future knowledge production, human-machine co-construction must ensure the subjectivity and innovativeness of humans, and at the same time, a comprehensive governance

基金项目：山东省社会科学规划研究重点项目“数智社会‘数字劳动’的资本逻辑批判与治理研究”（25BZXJ02）。

作者简介：丁大尉（1976—），男，山东烟台人，烟台大学文学与新闻传播学院教授，研究方向为科技政策与战略。

framework of integrating heterogeneous actors' participation and cognitive justice must be established to ensure that the technical logic of GAI conforms to the inherent needs of human society development.

Key words: GAI; knowledge production; AIGC; man-machine collaboration

自 1956 年达特茅斯会议提出"人工智能"(artificial intelligence, AI)概念以来,人工智能技术就已经对人类知识的生产方式产生了深刻且重大的影响。2010 年至今是人工智能深度学习飞速进步的黄金时期,已经基本解决了达特茅斯会议提出的最为困难的问题,机器已经能够以研究人员所希望的方式阅读、倾听和理解语言[1]。而生成式人工智能(generative artificial intelligence, GAI)的出现,已被学界和业界公认为人工智能技术的重大突破,代表了人工智能未来发展的新方向。GAI 一出世便展现出强大的知识生成甚至重建现有知识生产系统的能力。以语言互动类 GAI 为例,生成式 AI 通过建立大语言模型(large language models, LLMs)对大规模语料库进行预训练,从而使得模型具备丰富的语言知识。当下,AI for Science 以及应用于具体学科的 AI for Math 等大模型形式已从概念验证进入"垂直深耕 + 跨域融合"的爆发期。然而,GAI 的快速发展同样引发了社会各界对于 AI 驱动知识生产风险的广泛担忧。如何系统辨识 GAI 驱动知识生产带来的新机遇和新挑战,厘清其潜在的风险构型和伦理挑战,构建安全可控、富有活力的科研新生态,已成为学术界面临的重要课题。

当前学界对于 GAI 的研究主要集中在范式革命[2-3]、技术特质[4-5]、人机关系[6-7]、传播影响[8-9]等层面,鲜有系统性的风险反思及治理研究。然而,GAI 驱动知识生产是一个关涉科技与社会间复杂互动关系的系统工程,有必要进行深入完整的反思性研究。丰富和挖掘 GAI 驱动知识生产的内在逻辑及其可能带来的风险构型,对于正处于科技创新关键赶超阶段的我国来说,尤为重要且紧迫。作为 STS 研究的重要逻辑基底,技术—社会共构的分析方法已成为学界考察科技与社会间互动关系的重要方法论。本文拟基于技术—社会共构视角,探讨 GAI 驱动知识生产中的技术架构与社会权力结构间的复杂关系,也为进一步深入解决 AI 大模型知识生成中的人机共创、知识幻觉、知识孤岛等问题提供新的可能路径。

一、GAI 与知识生产的深度融合:以语言互动类大模型为例

大语言模型本质上是一种自动生成应答内容的机器创作方式,已经深刻变革了人类知识生产的传统方式,给当前学术界带来了颠覆性影响。"不同于过去的'分析式'AI(analytical AI),这类人工智能模型不仅可以通过对数据的学习来提炼信息、预测趋势,而且可以生成不同于学习样本的新内容"[10]。如今,GAI 已经可以独立完成诸如文献分析、数据处理、论文撰写等工作任务,这些任务在其原始训练数据或程序中并没有被明确指定。从这个意义上看,GAI 正不断扩展着知识生成、知识评价以及知识传播的边界,在重构知识生产新秩序的同时提供了更为个性化和交互性的知识服务,进而深刻影响着社会公众对于知识的获取和理解。

GAI 所建立的人机融合知识生产模式的方法论基础是人类与 AI 大模型深度协作下的科学发现与知识创新。它以海量语料数据作为基础,以大规模生成式预训练模型为支撑,通过反复的微调预训练确保人机对话的顺畅性。GAI 不仅可以进行检索文献资料、捕捉学术热点、提取关键信息等辅助性工作,还可以完成凝练学术问题、输出学术观点等创造性工作,甚至可以据此生成结构完整、逻辑合理的学术论文。以 ChatGPT 为例,超大型预训练语言模型 GPT - 3.5 在通过大量语料库反复训练的基础上,逐步学习并掌握了自然语言的基本规则和程式。在此基础上,再通过人类输入反馈的强化学习(RLHF)机制不断激活大模型的各项能力,并按照符合训练语料库中所内嵌的文化习惯、情感倾向甚至意识形态偏向的方式来理解和生成人类语言,并与人类互动。

GAI 还可以在现有知识储备的基础上进行自我理解能力的延伸拓展。具体来说,普通 AI 在解决问题时,主要依赖较为固定的算法规则,依据封闭的算法程序对输入数据进行拟合并输出结果,可以认为普通 AI 不具备自主思考能力。但 GAI 却与此不同,以 Gemini 多模态大模型为例,其语料

库中构建了较为完备的数理逻辑模型,并且依赖大模型强大的计算分析能力,形成了逻辑严密的思维链推导和理解过程,可以“提升论文深度和论证效力”[11]。另外,GAI 不仅在算法和算力方面,而且在判断准确性、价值选择以及逻辑推理等方面,都是普通 AI 无法比拟的。

另外,GAI 已经具备了较高的语言表达水平。一般来说,学术论文必须要以简洁清晰的话语体系将原理深奥、逻辑复杂的知识内容表达清楚。通过大语言模型和海量数据的预训练,尤其是在人类持续的强化学习和反复微调的基础上,GAI 已经具备了非常出色的语言组织和理解能力。AIGC 的文本风格已经十分自然流畅,与人类的对话也非常灵活自如,可以在对已有知识体系进行结构化处理的基础上,生成脉络清晰的知识图谱,甚至已经广泛应用于文学创作、新闻制作、医药医疗、线上教育等专业性很强的知识领域。

二、认知压迫:GAI 驱动知识生产的风险构型

GAI 驱动下的知识生产本质上是新技术逻辑与人类社会的共构。作为对人类认知的技术性模拟,GAI 在生成知识的同时也系统性复制并放大了人类社会既有的认知不公,并且伴随着知识异化、知识边界、知识权力化、知识语境集中等新的认知风险。由此,亟须定义人机共创的基本原则,构建多主体协同的综合治理机制,防范 GAI 可能引发的系统性风险。

第一,知识异化问题。GAI 将不断消解人类在知识生产中的主体性地位和创新能力。从社会层面看,基于大数据与大算力的数据理性正被过度信任 AI 的非理性所掩盖,人类甚至逐渐忘记了自身驱动知识生产的仪式感与独立性,“大模型技术既有改造未来世界、通向‘自由人的联合体’的可能性,也有可能在各种异己力量影响之下对人类社会产生负面的影响,并加剧异化的可能”[12]。对于生成式 AI 的过度技术崇拜无疑将陷入技术至上主义的泥沼,从而丧失人类特有的自主思考和独立判断能力。“随着生成式 AI 的进一步发展,对其过度依赖可能会将它视为知识权威和道德权威,进一步阻碍科学知识创新”[13]。这种情况甚至直接导致了社会层面对于未来 AI 的莫名恐惧,甚至“成为人们崇拜仪式下的力量核心,形成崇拜力量核心的符号资源圈层”[14]。

强替代认识论甚至把 AI 视为新的知识主体,人类反而退化为“AI 的工具人”。主体性的“外包”不仅会消解人类的深度思辨与原创能力,而且默会知识、直觉、情感也将被逐步弱化。所以,需谨慎对待人类认知惰性、思维创造消解等问题,防止 GAI 驱动知识生产中人类主体性的转移,“作为一种系统,不仅仅体现出‘自创生’的典型特征,在人与技术机器的深度融合中,更呈现出一种‘共创生’状态”[15]。另外,人类对于 AI 的价值不止是使用,其治理能力也同样体现在人类那些具有决策和导向意义的数据审查、清洗以及干预行为在 AI 知识生成中发挥着的基础性作用。

作为知识生产与信息传播的重要行动者,GAI 正越来越多地参与世界的诠释与建构,并将在未来世界的诠释权力中占据非常重要的位置。对于人类认知来说,这种机器知识将越来越成为一种异己的力量。然而,不同于人类思维逻辑在社会实践中的自然进化,AI 语言逻辑只是通过对高度复杂的神经网络设计去贴近人类的理解,所以必然带有天然的片面性和断裂性,“对它而言,输入和输出的信息只是单纯的符号串,它所处理的信息只是表面看来是信息”[16]。所以,AI 算法设计和数据分析中必须保持人类思维、情感和行为的独立性与独特性,使人类升级到更高阶的创造性、判断性和价值性环节。

第二,知识幻觉问题。GAI 驱动知识生产存在天然的局限性。除了算法本身的局限性外,良莠不齐的大数据网络信息甚至虚假信息也成为算法投喂的一部分。对于超大算力的数据吞吐能力来说,已经无从考究这些数据对于知识的应答机制产生的实质性影响。同时,AI 技术的反身性作用也在不断地对人类自身进行整体化规训,或者说其知识输出也在反向塑造人类的认知和理解方式,“技术人工物作为一种手段,使我们从以前无法经验的、不能感知的‘世界’通达到我们现在所在的领域”[17]。AI 教父辛顿曾警告,因为制造大量虚假文本、图片和影像,生成式人工智能将成为未来互联网中虚假信息的最强传播者。如今,GAI 生成的虚假信息更具迷惑性,大量“幻觉”导致人类辨识度更低、辨识成本更高。

AIGC 的知识生成是存在先天局限性的，其本身并不具备有效判断生成内容准确性、合规范性与合伦理性的能力。对于那些具有相当理论深度的学术问题，AI 更是无法提供真正有洞见的反思性见解。举例来说，用户的反复提问与确证对于 Gemini 理解用户指令来说，就是将分散的海量数据进行类人工的再次集中呈现，其生成结果可以多次反复滚入更大的数据背景中从而形成更为全面的回答。所以，GAI 很难生成具备重大创新意义的原创性发现，必须明确人类优于 AI 的情境理解和情感共鸣能力以及 AI 缺乏科学批判和反思能力的事实，认识到“人类在参与知识创造、科学决策、伦理判断和批判性思维方面的角色不可替代”[13]。

知识“幻觉”的风险逻辑源自技术因素与社会因素两个维度。就技术层面看，AI 大模型的技术架构与算法逻辑中内嵌着传统人际间的不平等认知权力关系。按照米兰达·弗里克关于认知不公的见证不公与解释学不公划分来看，大模型中的训练数据不公再现了人类社会中的证言不公，算法不公则对应着人类社会中的解释不公。从社会层面看，高技术企业、政府部门与社会公众间资源的不平等占有关系决定了认知权力的不断集中，这种集中在削弱个体认知自主性的同时甚至造成了整个社会公众的隐性认知依赖。

第三，知识权力化问题。相较于普通 AI，GAI 更加强化了生成性知识中的价值嵌入与权力渗透。尽管 AI 的知识储备已经远超人类个体，但算法学习与训练中一定包含训练数据的选择与清洗、计算模型的设计与实现等差异性环节，这就导致 AIGC 中往往包含强烈的社会价值偏好乃至意识形态偏向。比如，Claude 4（商用闭源大模型）算法训练中使用的主要是源自欧美主流国家体现西方价值观的数据集，“不管是其输出的观点，还是创造的秩序，当‘不透明’‘不公平’‘价值偏向’等不信任发生，相应的危机也就发生了”[18]。

还要看到，AI 机器端的技术黑箱意味着知识生成的权力被进一步集中强化，“合理”封装起来的算法逻辑也变得更加无从解构。科技企业与社会公众的不平等关系导致认知权力的不断集中，用户不自觉地形成错误的群体性认知，从而导致 GAI 大规模复制与强化既有权力体系中的不平等关系。所以，“明确 AI 与人类在知识生产中的责任，实现人机协同进化，可能是人类面对人工智能未来发展更为明智的选择”[19]。

另外，深度机器学习背后隐藏的算法偏见还将引发种族主义、性别歧视、年龄歧视等认知不公问题，倡导深度学习中的多元文化和社会价值观是确保 AI 知识公平正义的基础，“AI 决策的文化视角和底层算法的透明度对于技术的接受至关重要”[20]。这一方面要求在算法训练时保证数据的可靠性、可解释性和透明性，有效规避算法设计和数据选择中社会偏见隐匿的智能复制；另一方面需要采用内容识别、内容可追溯性等技术手段对知识内容进行评估和核查，以最大程度地排除知识生成中的潜在风险。

第四，知识语境集中化问题。GAI 将不断消解知识内容的多元性和情境性，形成所谓的“主流”知识霸权。人类投喂 GAI 的训练数据中包含明显的文化、地域以及种族差异。尤其是在全球化趋势日益明显的今天，基于人数占比形成的强势主流知识正被不断极化，地方性知识、差异性知识、边缘化知识等非主流知识因得不到足够的训练机会而走向式微。而且，符合强势文化的知识内容将不断强化用户接收相似信息时的价值认同，逐步扩大社会公众的认知偏见，从而形成歧视性思维的信息茧房。由此可见，看似价值中立的 AI 知识实际上正在潜移默化中规训着用户的认知和思考方向。

以 Claude 4 闭源大模型为例，目前接入其训练库的数据主要来自于互联网，符合主流价值观的通识性知识表述在网络平台中具有较高的传播力和影响力，这些数据对于 AI 大模型的投喂概率远超那些用户黏性较低的地方性以及非正式知识内容，这就导致地方性和非正式知识被 GAI 学习和呈现的机率很低。可以预见，就 GAI 这种生成性知识生产模式来说，很多地方性语言将因为在 GAI 大模型中得不到强化训练而走向消亡。除此之外，GAI 大模型很难理解特定历史和社会情境下的“意会”知识。当前的大语言训练模型给出的回答往往是把具体语境做了平均化处理，以至于在面对具象化的情景问题时，也往往给出非常宽泛

和抽象的答案。另外需要注意的是,GAI 还不能明确表达隐匿性和经验性知识,更不用说那些不能用语言系统直接表达的默会知识。斯科特在《国家的视角》中曾批判过“非语境化知识”,他认为经过批量标准化处理的知识往往显得过于抽象和简化,人们应该更多地关注“非正式知识、民间智慧和地方性实践知识”[21]。

三、人机共建:GAI 驱动知识生产的协同治理路径

随着以 GPT、AlphaFold 等为代表的 GAI 深度嵌入科研实践,知识生产正经历从专家主导向人机共建的范式转型。从技术—社会共构的视角看,关于 GAI 驱动知识生产,须构建融合认知正义的伦理规约、知识标准、社会评估三位一体的综合治理框架。政策制定者、技术企业、科学共同体以及社会公众等异质性行动者应共同开辟出一条负责任的、保证人类认知主体性的人机共建道路,明确 AI 参与知识生产的边界和有限性,在保证人类认知主体地位的同时,将认知平等和正义价值导向融入大模型的全生命周期。

第一,重构 GAI 知识生成中的“类主体性”属性及其责任伦理框架。AI 并不具备判断知识内容是否符合道德伦理规范的能力,如果其训练语料库受到“伦理污染”,比如受到测试者的持续恶意恐吓、不良诱导等,AI 仍会按照既定指令输出,无法从科技伦理层面对输出内容进行自我完善。大模型经常出现的事实性错误说明其生成内容并不可靠,原因是 AI 对于恶意投喂和带有明确目标指向的数据训练没有任何分辨能力,算法仍将机械地输出大量所谓“幻觉”信息。所以,GAI 开发中应贯彻价值敏感性设计(value sensitive design)理念,将人类核心价值嵌入技术设计环节,同时建立积极的人机交互反馈与识别机制。

就学术界来说,大量恶意生成的虚假论文和研究报告的传播将造成学术垃圾的泛滥,辨识成本极高,由此导致的抄袭、剽窃、造假等学术不端问题也将变得更为复杂。AIGC 的大量涌现还使得知识的创新评价工作变得更为困难,带来了诸如科学家如何规避 AI 陷阱、学术创新是否受到 GAI 大模型的影响等新的学术伦理问题。另外,信息泄露也是 AI 大模型知识生产的一大伦理隐患,“在训练过程中,模型可能会间接地‘记住’部分训练数据中的敏感信息,如个人隐私、商业机密、专利信息等”[22]。但现有伦理审查体系仍缺乏对 AI 生成内容的责任归属判定标准。

如果 AIGC 中包含错误信息,那么即使对相关内容进行完备的错误验证,也无法追溯触发 AI 错误内容的有效数据。正如有学者所言,“在缺乏对语言背后常识、因果关系和世界模型的认知时,基于大语言模型的人工智能系统可能会对人类社会带来前所未见的风险”[23]。因为现有版权体系中规定的产权人是自然人,显然 GAI 不具备独立承担侵权责任的能力,这就导致此类问题无法在现有知识产权政策框架内解决。同时,人类已经无法还原生成性知识的全部原始素材和推论方法,追溯生成性知识中的错误内容也变得更为困难。可见,AI 的“类主体性”造成了 AIGC 的责任转移与伦理张力,责任备案、过程标注以及使用披露机制等的缺失已经影响到了 GAI 知识生成中的透明性以及人机间的信任关系。

第二,重建人机共建模式下的知识合法性标准与知识的验证逻辑。GAI 的对话生成模式下,因为算法中大量非线性神经网络的关系是反复迭代的,所以根本无法追溯其输出内容的有效触发数据。AIGC 生成过程中涉及算法开发者、设备资源、数据拥有者、用户等众多行动者,这也同步增加了 AIGC 合法性验证的难度。当前,AI 的知识类训练语料主要来自数据库中的论文著作、机构知识库、个人学术博客、知识评论区等在线内容,其中很可能包含很多没经确证的内容,所以“需保证生成式人工智能在学术研究和出版领域被道德和负责任地使用”[24]。本质上看,GPT、AlphaFold 等大模型的知识生成主要基于高维数据的相关性关系,科学家根本无法全面理解这种机器黑箱逻辑。

如今,GAI 的知识生成速度已呈指数级增长,想要对 AIGC 的所有内容进行完备的知识验证几无可能。就人机新关系来讲,AI 大模型的封闭运算基本剥离了科学共同体对于知识的协商、解释以及批判的权力,其知识生产“本质上无法在创造力与自身约束之间取得平衡:它们要么过度生成

(产生真相或谎言,同时支持道德和不道德的决定),要么生成不足(对任何决策表现出不承诺和对后果的漠不关心)"[25],于是 GAI 生成内容是否合法就成为一个关键问题。*Science* 杂志已经明确禁止将 AI 列为作者,AI 生成的文本、图形、图像等内容也不允许在文中作为支撑材料使用[26]。另外,GAI 生成知识的表征方式不同于传统科学的透明性、再现性等要求,这也对现有科研规范体系形成了系统性挑战。

有学者指出,相对于一阶(原创性)知识和二阶(在整理加工一阶知识基础上形成的)知识来说,AI 的生成性知识可以作为"三阶知识"[27]。用于 GAI 模型训练的语料库主要来自网络上与知识内容相关的各种数据集,比如知识数据库、电子书籍、开放学术类网站(维基百科等)、机构仓储、政务网站、各种社交平台等,这些素材属于不同形式的一阶和二阶知识。本质上看,AI 大模型就是在对这些训练语料反复进行输入输出的模仿、推论和预测的基础上对已有知识内容进行再组织和再阐释,从而生成超越自然语言的多种新的知识形态,比如文本、音视频、程序代码等。所以,这种知识生成不同于一阶知识的原始创新,"ChatGPT 对我们知识的整理是一种物理集合,它能够整合知识和全网大数据,但它并没有让知识产生化学反应,它只是物理反应"[28]。由此可见,GAI 的知识生成更应该被理解为一种"操作性知识",其输出结果的有效性往往由技术性能定义,而非建立在理论解释或推理机制的基础上。

第三,建立完善 GAI 知识生成中的审查评估与政策回应机制。从知识论视角看,知识等同于"被确证的真信念",然而从技术层面看,GAI 主要是根据大模型进行数据训练时建立的词与词之间的关联性,按照最优的概率模型为前面的词匹配下面的词。AI 只是机械地输出物理形式上的"数据",这些数据汇集成的文本、声音、图像等生成式内容的知识属性仍需要人类来赋予,它们是否负载新知识更是需要人类的辨识和验证。换句话说,AI 只是在数据符号层面进行数字化处理,它本身并不具备理解并处理语法语义的功能。AI 不具备"对错""好坏"等价值判断能力,它甚至经常生成与已有知识体系以及当前社会共识等价值观不自洽甚至完全冲突的知识内容。

就目前的情况看,大模型的设计与发布主要是由平台完成的,科研机构基本无法介入其技术规范制定,政府的监管责任也没有贯穿 GAI 的全生命周期。其审查机制目前主要局限在前置审批的范围中,仍缺乏对于 AI 大模型动态演化的追踪和结果评估。与欧盟 AI 法案等国际先进经验相比,我国尚未建立面向 AI 系统的分级监管与试点沙盒制度。总之,必须在推动 GAI 技术创新发展的同时建立适合我国语境的综合性治理框架,以确保人机共建中人的主体性以及 GAI 的负责任发展。

关于 GAI 驱动知识生产的风险治理,未来急需做好两方面的工作。一是需要不断加强 AI 大模型透明性和可解释性的审查评估。AI 大模型的神经网络中包括数百亿级的参数集,应将更多科研机构、技术公司和用户引入计算模型,共同监督和影响其走向。目前的一个大胆尝试是基于大语言模型提供超然于既有意识形态的客观性参照,让 GAI 为用户提供多种观点输出,用户可以基于自身价值偏好请求获得同类信息,并与之形成友好互动。这种做法实现了部分的算法透明,用户可以在倾向的认知基模内理解问题。二是需要不断加强法律监管和问责机制,加强对于 AI 的法律监管和可能灾难性后果的防范,比如制定严格的数据隐私政策、规范数据的使用范畴、实施 GAI 的专有道德规范等。总之,关于 AI,需建立包括科学共同体、高新技术企业、科学组织、政府乃至社会公众等众多行动者在内的多元综合治理共同体,共同参与规制 GAI 的未来演化。

四、结语

本文基于技术—社会共构的视角对 GAI 驱动知识生产的风险问题进行了分析。除了作为语言互动的大模型外,AI for Science、AI for Social Science 以及应用于具体学科的 AI for Math 等 GAI 新兴赛道方兴未艾。2026 年 3 月 27 日的中关村论坛年会上,AI for Science 成果集中亮相,标志着"读—算—做"智能科研闭环即将走向规模化产业应用。如果未来 GAI 能够在生物医学、合成化学、航空航天等领域自主感知、决策甚至行动,那么整

个科研效率将获得指数级提升。然而,GAI 在重建知识生产、传播和应用秩序的同时,也内嵌着一系列新风险。GAI 生成性知识本质上是异质性行动者集体参与下的对于已有知识体系的解构和重建,政府部门、高新技术企业、科学共同体甚至社会公众等异质性行动者均应以不同方式集体参与这场知识革命,实现知识生成中的人机共建和协同治理。如何在促进 GAI 提高知识生产效能的同时,有效规避其潜在风险,已成为摆在我们面前的重要课题。这对于正处于以科技创新引领新质生产力发展与大国科技竞争叠加背景下的我国来说,意义尤为重大。

参考文献:

[1]DEAN J. A golden decade of deep learning: computing systems & applicatons[J]. Daedalus, 2022,151(2):58-74.

[2]张辉,刘鹏,姜钧译,曾熊. ChatGPT:从技术创新到范式革命[J]. 科学学研究,2023(12):2113-2121.

[3]HECHLER E, OBERHOFER M, SCHAECK T. Deploying AI in the enterprise: IT approaches for design, devops, governance, change management, blockchain, and quantum computing [M]. New York: Apress Media LLC., 2020:3-22.

[4]LANGENKAMP M, YUE D N. How open-source machine learning software shapes AI[C]// Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. New York:Association for Computing Machiery, 2022:385-395.

[5]MITCHELL M, KRAKAUER D C. The debate over understanding in AI's large language models[J]. Proc. Natl. Acad. Sci., 2023, 120(13):1-5.

[6]张铖,李正风,潜伟. 从 ChatGPT 到人机协作的知识共建[J]. 科学学研究,2023(12):2131-2137.

[7]KASNECI E, SESLER K, KUCHEMANN S, et al. ChatGPT for good? on opportunities and challenges of large language models for education[J]. Learning and individual differences, 2023, 103(1):1-13.

[8]王建磊,曹卉萌. ChatGPT 的传播特质、逻辑、范式[J]. 深圳大学学报(人文社会科学版),2023(2):144-152.

[9]PAVLIK J V. Collaborating with ChatGPT:considering the implications of generative artificial intelligence for journalism and media education[J]. Journalism and mass communication educator,2023,78(1):84-93.

[10]陈永伟. 超越 ChatGPT:生成式 AI 的机遇、风险和挑战[J]. 山东大学学报(哲学社会科学版),2023(3):127-143.

[11]姜涛,叶思雨. 生成式人工智能浪潮下的学术生态:机遇、挑战及应对[J]. 河北科技大学学报(社会科学版),2024(6):85-92.

[12]高奇琦. 共有、共治与共享:大模型技术的发展之路[J]. 华中科技大学学报(社会科学版),2024(2):57-65.

[13]郑泉. 生成式人工智能的知识生产与传播范式变革及应对[J]. 自然辩证法研究,2024(3):74-82.

[14]孙琦,陈力丹. 从力量崇拜到认识晕轮:ChatGPT 的传播仪式[J]. 传媒观察,2023(7):17-22.

[15]孙玮,李梦颖. "码之城":人与技术机器系统的共创生[J]. 探索与争鸣,2021(8):121-129.

[16]李大强,许双. ChatGPT 会说话吗:AI、心灵与语言[J]. 社会科学战线,2023(5):48-56.

[17]周瑄,戴瑞. 后现象学技术哲学视域下设计的认识论问题研究[J]. 湖南社会科学,2022(5):42-47.

[18]王建磊,曹卉萌. ChatGPT 的传播特质、逻辑、范式[J]. 深圳大学学报(人文社会科学版),2023(3):144-152.

[19]崔伟奇,程倩春. 从司马贺的"人工科学"观看人工智能的哲学基础[J]. 自然辩证法研究,2023(11):39-47.

[20]YOGESH K. International journal of information management [EB/OL]. (2019-08-02)[2026-01-16]. https://doi.org/10.1016/j.ijinfomgt.

[21]伯克. 什么是知识史[M]. 章可,译. 北京:北京大学出版社,2023:157.

[22]刘智锋,吴亚平,王继民. 人工智能生成内容技术对知识生产与传播的影响[J]. 情报杂志,2023(7):123-130.

[23]MITCHELL M, KRAKAUER D C. The debate over understanding in AI's large language models[J]. Proceedings of the national academy of sciences of the United States of America, 2023, 120(13):1-5.

[24]LUND B D, WANG T, MANNURU N R, et al. ChatGPT and a new academic reality:artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing[J]. Journal of the association for information science and technology,2023, 74(5):570-581.

[25]Chomsky N. The false promise of ChatGPT[EB/OL]. [2026-03-06]. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

[26]THORP H H. ChatGPT is fun, but not an author [J]. Science, 2023(6630):313.

[27]肖峰. 生成式人工智能与知识生产新形态[J]. 学术研究,2023(10):50-57.

[28]张洪忠,黄烈民. ChatGPT 的技术逻辑、社会影响与传播学未来[J]. 江西师范大学学报(哲学社会科学版),2023(2):25-31.