聚知刊出版社
JZK publishing

人工智能工具视域下高校学术诚信建设与应对策略

和献歌

南通理工学院，江苏南通，226001；

摘要：人工智能技术在高校学习与科研中广泛应用，但部分学生对其机制与局限认知不足，存在过度依赖、直接使用生成内容而不核验的问题，易引发学术诚信风险，影响研究质量。高校应加强引导，规范学生合理使用人工智能工具，坚守学术诚信与独立思考，保障学术生态健康发展。针对人工智能工具滥用现象，高校应采取及时有效的应对措施。引导学生正确识用人工智能工具，坚守学术诚信，保持独立思考与创新的能力。

关键词：人工智能工具；高校学生；学术诚信；应对措施

DOI：10.69979/3029-2735.26.01.100

引言

人工智能技术已广泛渗透于各领域，高等教育领域亦受其深刻影响。作为知识探索与创新实践的主体，高校学生日益倾向于借助人工智能工具辅助学习与科研，这类工具可快速检索资料、生成规范文本，有效提升学习科研效率。但正如所有新兴技术均具双面性，人工智能工具在带来便利与机遇的同时，也引发了一系列问题与挑战。

一方面，多数学生对人工智能工具的运行机制与内在局限性认知不足，甚至将其简化为单纯的信息检索与内容生成工具，部分学生直接利用其生成学术论文、查询数据，却未进行必要的核验与补充。这种过度依赖与滥用行为，不仅会降低学术成果的真实性与可靠性，更会损害学术诚信，违背学术研究的核心原则。另一方面，基于大语言模型的人工智能工具虽功能强大，却存在“幻觉”等固有缺陷，易生成不实信息，这一特性也进一步加剧了学术风险。因此，高校学生需明晰人工智能工具的运行逻辑与局限，以辩证思维合理运用，保障学术研究质量。

本文旨在探讨高校学生使用人工智能工具面临的现实挑战，分析其滥用现象对学术生态、教育质量及个人发展的多重影响，并提出针对性应对措施。研究以期为高校学生提供科学的使用指导，助力其在享受技术便利的同时，坚守学术诚信、保持独立思考能力，实现学术道路的稳健发展。

1 高校学生人工智能工具滥用的现象与影响

高校学生人工智能工具滥用主要表现为作业/论文代写、考试作弊、科研造假三类学术不端行为：学术写作中，学生在各类作业中使用人工智能工具却不披露，或借助人工智能工具改写降重、付费购买人工智能生成的定制化内容；科研中，通过 Consensus 等工具自动生成缺乏原创性的文献综述，或利用人工智能修改、伪造数据，影响学术成果真实性；考试中，借助语音助手、图像识别等人工智能技术作弊，甚至出现换脸、语音模拟替考现象。

人工智能工具学术造假会造成虚假数据泛滥。ChatGPT 类生成式人工智能模型的主要目标是模拟人类语言，而不是提供一个准确的回答。因此，为了获得用户的认可，它往往倾向于优先考虑编造虚假信息来提供一个清楚的回应而不是精确的答案。[1]人工智能工具可快速生成看似合理却虚假的实验数据与研究成果，污染学术期刊与会议，削弱科研可信度。大语言模型能写出逻辑严密、难被查重与传统评审识别的论文，加剧“出版即正确”的错觉。人工智能 造假违背科研可重复性，误导后续研究，浪费人力与资金；相关事件会助长反智主义，降低公众对科研的信任，破坏学术公平与严谨氛围。同时，虚假论文会污染训练数据，形成 “垃圾进、垃圾出” 的恶性循环，反噬大语言模型本身。AIGC 在提升科研效率的同时，也带来抄袭、虚假成果等学术失信问题，引发学界对作者身份、知识产权与研究透明度的担忧。

高校一贯高度重视学术诚信建设，而人工智能的滥用行为正在动摇这一重要基础。学生若过度依赖人工智能，容易丧失独立思考与自主解决问题的能力，直接影响高校人才培养质量。长期依赖人工智能工具，还可能导致学生知识掌握不扎实、综合能力发展滞后，进而制约其未来职业竞争力。当前主流剽窃检测工具对人工智能生成内容的识别能力有限，使得学术不端行为更难被有效甄别。不少高校仍主要依靠教师经验判断人工智能

滥用情况，这大幅增加了教师的工作量，挤占了原本用于教学与指导的时间。此外，学生过度依赖人工智能完成学习任务，会减少师生、同学之间的真实交流与互动，弱化教育过程中的人文关怀与情感传递。

2 高校人工智能工具滥用的应对措施

2.1 强化认知教育，明晰工具边界

面向学生开展人工智能应用规范培训，讲解工具原理、局限与风险，破除 “AI 生成即正确” 误区，树立技术为辅助、思考为核心的科研理念。

目前大多数人工智能工具都是基于大语言模型（LLM）开发的。当前最先进的生成式 LLM 如 GPT-4，Llama2[2]等基于海量人类生成的语言训练而来，以拟合人类语言分布特性为优化目标，生成的语言在语法规则、逻辑结构、流畅程度等方面都已达到甚至超过了普通人的文本生成水平[3]。开发者可直接使用预训练模型，或在其基础上进行微调，从而快速搭建各类人工智能应用。这些应用依托大语言模型强大的语言理解与生成能力，结合其他技术手段，形成了丰富多样的人工智能产品。例如 ChatGPT、文心一言等自然语言处理工具，可广泛应用于文本生成、智能对话、机器翻译等任务。当前人工智能工具还同时具备多模态能力，能够同时处理图像、音频、视频等多种类型数据。这种多模态融合可以输出更全面、更精准的结果，适用于更复杂的应用场景。如 Deep Seek、NEXUS-O、DALL - E 等，通过将大语言模型与计算机视觉模型相结合，实现了文本、图像、音频等多模态信息的统一处理。LLM 已经成为许多人工智能工具的核心基础，通过与其他技术深度融合，相关工具得以实现多样化、高效化的人工智能应用。

尽管 LLM 具备很强的学习和创作等通用能力,但受限于其黑箱模型特性和领域知识缺乏等原因,其可解释性、可信赖性和领域能力都还难以令人满意[4]。虽然这些模型在处理自然语言任务方面仍然存在一些局限性和潜在问题，可能导致生成的文本包含错误、虚假信息或不符合现实的内容。LLM 是基于统计学习的，它们通过分析大量的文本数据来学习语言模式和结构。生成文本时，模型会根据输入的上下文和已学习的概率分布来预测下一个词或句子。因此，生成的内容是基于数据中出现的模式，而不是基于事实的验证。模型在生成文本时，通常没有直接访问外部知识库或实时数据的能力。这意味着它们无法验证生成内容的真实性，只能依赖于训练数据中的信息。如果训练数据中存在错误或偏差，模型可能会学习到这些错误并将其反映在生成的文本中。此外，如果数据不够多样化，模型可能无法处理某些特定领域或主题的问题。模型的训练目标通常是最大化生成文本的流畅性和连贯性，而不是确保内容的真实性。这种优化目标可能导致模型在生成文本时更关注语言的合理性，而不是事实的准确性。模型的输出高度依赖于输入的上下文和提示。如果输入的上下文存在错误或误导性信息，模型可能会生成与之相符的错误内容。例如，在回答关于历史事件、科学事实或当前事件的问题时，模型可能会提供错误的细节或完全错误的信息。生成的文本可能在逻辑上不一致或存在矛盾。模型可能无法完全理解复杂的因果关系或逻辑推理。

当数据集本身呈现出偏见时，由此衍生出的结果一定存在某种偏见。 ChatGPT 等国外 LLM 主要以英语文本为训练数据，其生成内容具有狭隘的以西方，甚至美国为中心的论调。例如，在涉及性别、种族或文化的问题上，模型可能会生成带有偏见的内容。生成式人工智能的持续应用将导致大量人工智能合成数据进入互联网，这些合成数据很可能会被再次投入机器学习，导致虚假数据的恶性迭代循环[5]。偏见审查是现有政策中明显缺位。这种关注不足可能使更多偏见内容渗入学术成果，融入语言模型数据库中，损害研究的客观性与公平性。[6]

2.2 加强技术监管，提升甄别能力

人工智能软件生成类似于人类创作的结构化新内容，并非简单的文本复制，因此很难区分人类创作的内容与 ChatGPT 生成的内容[7]。有学者通过研究已验证，“传统的基于文字相似性的学术不端检测系统对于（AIGC）的文本无能为力”。[8]高校应积极引入或研发先进的人工智能生成内容识别技术，构建专业的人工智能检测系统。例如，南开大学 PaperMate 智能检测系统 是高校自主研发、面向学术场景的 人工智能 生成内容检测与科研辅助平台，核心解决人工智能 文本识别不准、误报漏报高、检测效率低等痛点，可直接作为高校构建人工智能 检测体系的典型案例。透明度义务是新治理规则的基础，是人工智能的设计者、开发者、使用者如实记录或披露设计、开发、使用人工智能具体情况的义务[9] 。透明度义务是学术诚信的基石，是“看得见的诚信”。作者履行透明度义务的核心内容是履行“归认来源”义务，[10]部分高校和教师尝试使用人工智能检测工具（如 ChatGPTzero、Turnitin）来识别人工智能生成内容，并要求学生如实披露 人工智能使用情况。这也是在 ChatGPT 介入学术论文创作的场景中，对

论文作者的创造性和学术贡献进行鉴定和评价的重要依据。如实记录和公开 AIGC 的使用情况，是 AIGC 时代履行透明度义务的具体体现。多所高校已发布人工智能使用规范，明确限制人工智能在学术作业中的使用比例。例如，上海交通大学发布了《规范学生使用人工智能工具的教师指南》，以注释、说明、解释的方式明确 GAI 使用者透明度义务的范例。

2.3 改革教学评价

2.3.1 强化过程性评估

高校应强化过程性评估，构建多样化考核体系与持续反馈机制，弱化传统终结性评价带来的局限性。通过减少单一标准化、记忆型考核，采用课堂限时写作、现场口试、实践操作、小组研讨、阶段性项目报告等多元化考核方式，能够有效降低人工智能替代空间，更真实、全面地反映学生的知识掌握、思维过程与综合能力。以随堂小测、课堂互动表现、中期成果汇报、阶段性任务验收等方式替代期末 “一考定终身” 的评价模式，有助于动态追踪学习效果，及时发现并纠正学习偏差。同时，建立全过程学习档案，记录并收录学生的课程笔记、实验记录、项目迭代过程、阶段性反思与总结日志等材料，从结果评价转向过程评价 + 成长评价，让学习轨迹可追溯、可观测、可评价。

2.3.2 转向高阶能力评估

在考核内容上，应转向高阶能力导向评价，减少对低阶重复性知识与标准化答案的考察，更加聚焦学生的批判性思维、创新意识、问题解决、自主探究等高阶能力培养。围绕专业核心素养设定明确的能力评价指标，如逻辑分析、数据处理、团队协作、学术表达与实践创新等，而非单纯考查知识记忆与公式复述。例如，工科专业可重点考核方案设计、系统搭建、工程实践等能力，文科专业可强化论证分析、观点建构、现实解释等能力。通过案例分析、情景模拟、真实课题研究、校企合作项目、社会调研实践等贴近现实场景的任务，让学生在解决复杂真实问题的过程中展示综合素养。教师可从研究设计、方案执行、团队协作、成果展示与反思改进等多个维度进行综合评定，真正实现 “以评促学、以评促教”，推动人才培养从知识记忆向能力提升转型。

3 结语

人工智能在带来便利的同时，也伴随着诸多挑战。无论是技术本身的局限、滥用引发的学术诚信问题，还是对教育质量带来的潜在风险，都提醒我们在享受技术红利的同时，必须保持理性与审慎。高校学生作为知识学习者与未来创新者，既要善用 人工智能 提升学习与研究效率，也要警惕技术依赖与潜在风险。高校则应承担起引导责任，通过完善 人工智能 使用规范、改革评价考核机制、加强学术诚信教育，帮助学生树立正确的技术观与学术观。在技术应用与学术规范之间寻求平衡，是新时代高等教育面临的重要课题。这需要学校、教师与学生共同努力，唯有坚持合理、规范、负责任地使用人工智能，才能守住教育质量与学术底线，培养出兼具创新能力与诚信素养的高素质人才。

参考文献

[1]刘辉，雷崎山. 生成式人工智能的数据风险及其法律规制 [J]. 重庆邮电大学学报(社会科学版) ，2024，36（04）：40-51.
[2]TOUVRON H, MARTIN L, STONE K, et al. Llama 2: Open foundation and fine-tuned chat models [EB/OL].(2023-7-19)[2025-9-5].https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/.
[3]姜毅，杨勇，印佳丽，等. 大语言模型安全与隐私风险综述[J]. 计算机研究与发展,2025，62（08）：1979-2018.
[4]PAN S,LUO L,WANG Y,et al.Unifying large language models and knowledge graphs:A Roadmap [J]. IEEE Transactions on Knowledge and Data Engineering,2024,36(7):1-20.
[5]游俊哲. ChatGPT 类生成式人工智能在科研场景中的应用风险与控制措施 [J]. 情报理论与实践，2023，46（06）：24-32.
[6]徐拥军，卢思佳，殷名. 国内外学术出版机构 AIGC 使用政策研究[J]. 图书馆建设，2025（04）：11-20.
[7]BISHOP L.A computer wrote this paper: What ChatGPT means for education,research,and writing [J].SSRN Electronic Journal,2023.
[8]沈锡宾，王立磊. 人工智能生成学术期刊文本的检测研究 [J]. 科技与出版，2023（08）：56－62.
[9]程睿.ChatGPT 介入学术论文透明度义务的可视正义功能[J].科学学研究，2023，41（12）：2138-2146.
[10]方流芳. 学术剽窃和法律内外的对策 [J]. 中国法学，2006（5）：155－169.

作者简介：和献歌（1991—），女，河南新乡人，助教，硕士。研究方向：教育学、翻译学。