ISSN：3078-8994
EISSN：3104-4492

大学生生成式 AI 使用现状、风险及治理平台构建

何泺为 贺梓恒

宜宾学院法学与公共管理学院，四川 宜宾 644000

摘 要：基于四川省高校 1640 份有效问卷及 13912 条网络评论，分析大学生生成式人工智能的使用现状、学习效益与潜在风险。研究发现，生成式人工智能广泛应用于课程学习与资料查阅，在提升学习效率的同时，也引发认知依赖、独立思考弱化及学术诚信等方面的担忧。据此，提出分级引导策略，设计集行为监测、风险提示、信息核验与素养培养于一体的教育治理平台，以促进技术应用与自主学习能力的协调发展。

关键词：生成式 AI；大学生；学习风险；平台构建

DOI：10.64649/yh.shfzykjcx.issn3078-8994.202610006

0 引言

生成式人工智能能够依照自然语言指令生成文本、代码与图像，并在连续对话中解释概念、修改表达和提出方案。与传统搜索工具相比，它不仅提供信息入口，还可以直接形成看似完整的答案。技术降低了操作门槛，也带来“由谁完成思考”的问题。查找资料与直接提交生成论文，涉及截然不同的学习过程。关于生成式人工智能的教育应用，既有讨论通常强调效率提升、个性化支持等机会，也关注内容失实、认知依赖与学术不端等风险 [1] 24。然而，对具体使用行为的判断不能止于“使用”或“不使用”的二分：同样的对话可能帮助理解，也可能跳过理解。单纯禁用或限时难以识别这种差别。

本文以大学生创新训练项目申报书汇集的调查与网络文本分析为主要材料。虽然项目原题涉及“青少年”，现有问卷的实际调查对象为四川省大学生，本文因而将讨论范围限定于高校学习场景。研究在梳理使用现状的基础上，分析技术支持与学习主体性之间的关系，并将分析转化为可供后续研发和检验的平台设计。平台仍属设计方案，风险阈值与干预效果尚待验证。

1 大学生生成式人工智能使用的现实图景

1.1 数据来源与观察范围

申报书汇集了两类材料：其一是四川省大学生使用生成式人工智能的 1640 份有效问卷，涵盖性别、年级、专业、应用场景、使用态度和风险感知；其二是围绕生成式人工智能话题采集的 13912 条网络评论，其中微博 1034 条、抖音 1788 条、哔哩哔哩 11090 条，经去重和无效信息筛除后保留 6643 条，有效率为 47.75%。网络文本采用分词、共词和情感分析。两类材料的观察对象并不相同：问卷反映受访大学生的自我报告，网络评论反映相关话题的公共讨论，后者不能被直接视为大学生使用情况的代表性样本。

样本结构限制了结论范围。问卷中女性占 62.3%，大一和大二学生分别占 30.1% 和 48.6%；专业分布以工学和理学为主，工学受访者为 963 人，理学受访者为 270 人。这些比例不宜推广至全省所有大学生，更不能推广至中小学生。未取得原始答卷与量表检验结果，本文仅讨论描述性发现。

1.2 从了解技术到嵌入学习

调查显示，1450 名受访者表示了解生成式人工智能，占有效问卷的约 88.4%；社交媒体是最主要的认知渠道，占比约 63.4%，课程学习与朋友推荐则构成其他重要来源。不少学生已能主动选择使用场景。不过，知晓工具名称并不等于理解其输出原理，更不能据此认定具有充分的事实核查能力。

在课程场景中，79.9% 的受访者表示使用过生成式人工智能辅助完成作业，76.8% 用于查阅课程相关内容，50.6% 用于处理课堂提问。在科研相关活动中，63.4% 用于辅助选择研究问题，53.7% 用于外文文献翻译。这些比例说明用途覆盖情况，而非实际使用时长。人工智能已成为随时可调用的工具，主要用于资料整合、思路启发和语言转换。

1.3 便利体验与风险担忧并存

受访者中，81.5% 对生成式人工智能持积极态度；83.6% 认为它能够节省时间，83.6% 认为它提供了新的思路，78.1% 认为它有助于理解复杂问题。与此同时，61.59% 担心工具影响独立思考，71.2% 将减少独立思考列为负面影响，72.6% 注意到可能出现的抄袭风险，80.1% 把较强依赖性视为一项风险。尤其需要指出，这些指标大多是态度或风险判断，不能直接等同于相应不良行为已经发生，更不能据此推断人工智能使用导致了认知能力下降。

ISSN：3078-8994
EISSN：3104-4492

网络文本分析从另一侧面展示了这种复杂态度。高频词包括“人工智能”“大模型”“训练”“数据”“工作”等，相关讨论既涉及技术进步和应用前景，也涉及就业与安全顾虑。两者可以参照，却不能合并计算。关键是如何划定便利使用与个人学习责任的界限。

2 生成式人工智能学习效益与风险的形成机制

2.1 认知支持与认知替代的分界

学习需经历提问、查证、论证与检验等环节。生成式人工智能能将分散信息组织为可读答案，为基础薄弱或缺少即时辅导的学生提供额外支持。当工具帮助解释术语、提出检索线索或反馈推理漏洞时，学生仍需自行完成理解和判断，这是一种认知支持。相反，当学生直接接受生成答案、减少原始材料阅读，甚至不经核实便用于考试或论文时，工具便可能从辅助者转变为认知任务的替代者。两者的差别并非输出形式，而是关键学习步骤是否仍由学生完成。[2]105

从认知负荷角度看，工具带来的效果也具有情境性。概念解释可降低初学者的信息搜寻成本，大量生成答案却也可能占用核验精力。更值得警惕的是“轻松得到答案”与“真正掌握知识”之间的错位：完成任务所耗时间减少，并不必然意味着知识迁移、独立论证和长期记忆同步改善。因此，评价人工智能的教育价值，不能只观察作业完成速度，还应考察学生能否独立复述推理过程、指出答案缺陷并在脱离工具后解决相近问题。

2.2 依赖风险为何不宜只用时长衡量

使用频率较高可能意味着学习任务多、技术接受程度高，也可能意味着对现成答案形成习惯，单凭使用时长难以区分。申报书提出探索日均使用 1.5 小时等风险阈值，但目前提供的问卷与评论资料不足以验证该数值具有普遍适用性。将它直接设置为限制标准，可能使认真核验资料的学生被误判，而少量使用却直接代写作业的行为反而被遗漏。[3]8 较为妥当的做法，是将行为频率与任务性质、主动检索、追问质量、事实核验及独立完成情况结合起来评价。

人工智能还可能改变学生的学习观念。如果课程仅以提交成品为考核目标，学生就容易把获得合格文本视为完成学习；若课堂要求展示证据选择、论证修改和错误纠正的过程，人工智能则更可能成为反思工具。这说明技术依赖并非孤立的个人习惯问题，也与课程任务、评价制度以及教师提供的指导方式有关。治理不能把所有责任转嫁给学生，更应反思现有教学环节是否为独立思考保留了空间。

2.3 学术诚信与信息可信的双重要求

生成式人工智能能够以流畅语言给出不存在的文献、未经验证的数据或似是而非的法律解释。学生如果缺乏核查意识，可能在无意间传播错误；如果明知课程要求独立完成，仍将模型生成内容冒充本人作品，则可能触及学术诚信规则。二者需要不同回应：前者侧重信息识别与核验训练，后者需要事先明确允许使用的范围、披露要求和评价后果。将“使用过人工智能”直接等同于“学术不端”，既难以操作，也不利于建立合理的人机协作规范。

此外，学习记录、提示词和上传资料可能包含身份信息、学习状况及未公开内容[4]35。无论平台由学校还是企业提供，行为分析都应遵循目的限定、必要性和最小化原则；学生应知道收集了什么信息、信息用于何处以及怎样退出相关分析。教育治理的目标是改善学习，不应因追求精确监测而形成持续、隐蔽的个人画像。

3 面向自主学习能力培养的治理平台设计

3.1 由“统一管控”转向“分级引导”

针对上述问题，平台设计首先需要确定治理对象。可将应用情境区分为三个层次：概念解释、资料线索与语言辅助等通常允许使用的情境；课程作业构思、程序调试和论文初稿修改等需要说明人工智能参与程度的情境；闭卷考核、明确禁止外部协助的作业以及伪造数据文献等应予限制的情境。区分依据应由课程教师事先公布，平台提供提示与记录功能，而不是独自决定学生是否违规。

在干预强度上，可以形成“反馈—提醒—转介”的渐进机制。一般使用提供每周用途概览和自我评价工具；出现频繁索取完整答案、缺少核验等迹象时，推送反思性提示和学习建议；当学生主动寻求帮助或遇到明显的规则冲突时，再将问题转交教师进行解释。

3.2 以四项功能回应具体学习问题

第一，行为反馈模块。学生在知情同意下自主记录工具使用目的、学习任务和时间，平台采用统计图呈现不同学科与任务的使用结构。反馈重点不仅是时长，还包括“使用前是否独立列出问题”“是否复核关键事实”“是否能解释最终答案”等自评项目，帮助学生发现自身学习习惯。原申报书提出的分学科环状统计界面可以作为这一模块的可视化起点。

第二，问题引导模块。当学生输入“直接帮我完成整篇报告”等请求时，平台不必机械

ISSN：3078-8994
EISSN：3104-4492

拒绝，而可优先询问题目要求、已有思路和拟使用的证据，提供提纲检查、反例讨论或推理提示。若属于明确禁止人工智能协助的考核任务，则提示遵守课程规则。平台不应仅凭敏感词给学生贴上“依赖”标签。

第三，信息核验模块。平台为模型输出附加可核查清单，提示识别具体事实、数字、引文与来源；对于未能确认的内容明确标识“待核实”，引导学生查阅教材、原始论文或权威数据库。技术设计应避免声称可以自动保证答案真实，算法评分不能代替事实认定。核验记录可服务于教学反馈，帮助教师观察学生是否具备基本证据意识。

第四，素养训练模块。围绕提示词表达、信息检索、模型局限、学术引用和个人信息保护设置短任务。例如，要求学生比较人工智能答案与教材解释的差异，为一段生成式论证找出能够支持或反驳它的证据，或者说明哪些个人材料不应上传。训练反馈用于改善学习策略，而非额外排名。四个模块相互衔接，才能把单纯的“行为管理”转化为可感知的能力培养。

3.3 明确实施边界和评价办法

平台宜先以自愿参与的课程试点验证可用性，采用匿名化或去标识化的汇总数据观察不同任务中的使用方式。学校需要明确数据保存期限、访问权限、学生更正与撤回渠道；教师原则上只能获得教学所必需的反馈，不宜默认查看全部对话。对可能涉及个人敏感信息的内容，应避免采集或另行进行合规评估。特别是眼动追踪等实验性方法，应在充分知情和伦理审查条件下独立开展，不能直接作为平台常态化监控手段。

3.4 学校、教师与学生的共同责任

平台仅是教育治理的辅助设施。学校应制定清楚、按课程类型区分的人工智能使用规范，明确可用范围和披露方式；教师需要调整作业设计，增加草稿、口头答辩、资料批注和修改说明等过程性要求；学生则应对提交作品的事实、引用和最终论证承担责任。技术开发者负责提供合理的隐私保护和纠错渠道。责任明晰，才能避免平台沦为数字化监督工具。

4 结语

基于四川省高校 1640 份问卷及相关网络评论材料，可以看到大学生对生成式人工智能存在鲜明的双重评价：它能改善信息获取和任务组织体验，也带来对独立思考、内容可信和学术诚信的担忧。现有资料主要提供描述性证据，尚不足以认定特定使用时长必然导致能力衰退，也不足以证明某一干预模型已经有效。相较于用频率或时长简单区分“合理使用”与“过度依赖”，更值得关注的是学生是否保留了提出问题、核查证据和形成判断的主动权。

据此，本文提出以学习任务为单位的分级引导方案，并从行为反馈、问题引导、信息核验和素养训练四方面勾勒教育支持平台。其核心不是替学生选择答案，而是帮助学生认识哪些环节可以借助人工智能、哪些环节必须由自己完成。下一步研究仍需补充原始问卷的测量检验、访谈与纵向跟踪，并通过课程试点比较平台实施前后的实际表现。只有将可验证的学习效果与必要的权利保障结合起来，生成式人工智能才能更稳定地成为促进大学生学习的工具，而非取代学习过程的捷径。

参考文献：

[1] 卢宇 , 余京蕾 , 陈鹏鹤 , 李沐云 . 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例 [J]. 中国远程教育 , 2023, 43(04): 24-31, 51.

[2] 蒲清平 , 向往 . 生成式人工智能——ChatGPT 的变革影响、风险挑战及应对策略 [J]. 重庆大学学报 (社会科学版), 2023, 29(03): 102-114.

[3] 王闻萱 , 王丹 . 中国生成式人工智能 DeepSeek 的核心特征、价值超越及未来路向 [J]. 统一战线学研究 ,2025,(02):1-14.

[4] 柴莞庚 . 生成式人工智能对个人信息保护的挑战和应对 [D]. 北京外国语大学 ,2024.

作者简介：何泺为（2006.10—），男，汉族，四川泸州人，本科生，研究方向：教育学、知识产权。
通讯作者：贺梓恒（1993.08—），男，汉族，重庆人，法学博士，讲师，研究方向：法理学。
项目信息：省级大学生创新创业训练计划支持项目“现代教育背景下青少年对生成式 AI 的使用现状研究及管控平台构建”（项目信息：S202610641176）。