大学生生成式人工智能风险意识调研

陈晨 蒋广学

北京大学网络安全和信息化委员会办公室

近年来，在大学生群体中生成式人工智能较为普及。借助生成式人工智能工具，学生可以生成高质量的论文初稿，进行作业指导，甚至编写代码，其应用不仅提高了学习效率，还为学生的创作与思维拓展了新的可能性。然而，生成式人工智能也存在潜在风险，例如隐私窃取、恶意诱导、歧视不公、伦理道德等问题。坚持合理、规范和安全地使用生成式人工智能，是大学生使用这一技术的重要原则，也是当前高校教育发展面临的关键举措。

本文通过对北京大学学生使用生成式人工智能工具的调查，分析该技术在大学生群体中的实际应用现状，说明大学生生成式人工智能使用风险管理方面存在的问题和需求，以期为生成式人工智能融入高等教育提供安全管理相关建议。

一、调查对象及方法

本研究的调查对象是北京大学当前在读学生，包括本科生、硕士研究生和博士研究生，涵盖了理学、信息与工程学、人文学、社会科学、经济与管理学、医学等专业类别，调查结果较具有代表性和真实性。

本文主要采用问卷调查法，编制了“北京大学学生生成式人工智能使用情况调查问卷”，其基本构成包括学生基本

信息、对生成式人工智能的了解和态度、对生成式人工智能风险的认知和态度等部分。

本研究通过公共在线问卷平台“问卷星”发放，并借由北京大学网信工作系统的各部门、院系和单位同事的支持完成。调查问卷回收时间为 2024 年 6 月 18 日至 24 日，共回收 547 份有效问卷,样本分布涵盖不同学习阶段和专业方向。

二、调查结果及分析

（一）大学生了解生成式人工智能的基本情况

调查结果显示: 绝大多数大学生（87.2%）对生成式人工智能工具有一定程度的认知，反映了其在大学生群体中的普及程度较高。然而，仍有 12.8%的学生对这些工具的了解有限甚至完全不了解，说明不同学生在认知水平上存在一定差距。此外，94.1%的大学生已经使用过生成式人工智能工具，高使用率要求高校思考如何更好地利用这一技术促进学习、研究和创新。

（二）大学生使用生成式人工智能的情况

为了具体分析大学生使用生成式人工智能的相关情况，本文对使用频率、使用场景进行调查。结果显示，每周使用时长超过 3 小时的大学生占比 29.7%，小于等于 3 小时的大学生占比 60.3%。这表明当前生成式人工智能工具在大学生群体中已具有一定普及度，不过尚未达到广泛深入的程度。

针对使用过生成式人工智能的大学生进行调查发现，分别有 37.9%和 39.8%的大学生认为生成式人工智能“非常重

要”和“比较重要”。这表明，生成式人工智能已经在当代大学生群体中产生了较大影响，并逐渐被广泛接受和认可。这一调查结果不仅揭示了生成式人工智能技术在高等教育领域的快速渗透力，还预示了未来教育形态与学习方式可能发生的重大变革。

对使用频率和重要程度的交叉分析结果见表 5 所示，两者呈显著正相关关系。具体而言，每周使用时长越长，大学生群体对生成式人工智能的需求也越强烈。高频使用者更倾向于将生成式人工智能视为日常学习生活的重要工具，反映出其在学生群体中的实用性和依赖性正在逐步增强。与此同时，低频使用者中也仅有极少数人认为生成式人工智能不重要，表明即便是偶尔使用的学生，也大多认可其在学习和生活中的价值。

大学生生成式人工智能的使用场景调查采用多选的形式开展，见表 4 所示。结果显示，大学生在“学术研究”“课程作业”和“编程或代码辅助”场景使用生成式人工智能均超过 60%，其他场景相对较少。这说明当前生成式人工智能在大学生的学习和科研中发挥了更加重要的作用，也进一步印证了生成式人工智能在高等教育中的影响力和实际应用价值。

（三）大学生对生成式人工智能安全问题的认知

虽然大多数大学生认同了生成式人工智能的重要性，但大学生对于生成式人工智能安全问题的了解情况不容乐观，

仅有15.1%的大学生认为自己“非常了解”，36.9%的大学生“比较了解”，见表6。

通过对所在专业大类与对生成式人工智能安全问题的了解程度两个变量进行深入的交叉分析，发现不同专业大类大学生在生成式人工智能安全问题了解程度上存在差异，如图 1。信息与工程学专业和理学专业的大学生对生成式人工智能安全问题了解更多，这可能源于他们日常学习中频繁接触到相关课程、研究项目及实践机会。

通过对不同学段大学生对生成式人工智能安全问题的认知程度进行深入的交叉分析，发现本科生、硕士研究生、博士研究生在该问题上的认知差异，如图 2。博士研究生对生成式人工智能安全问题的了解程度高于硕士研究生，硕士研究生则优于本科生。

根据文献[1]和文献[2]，本文将生成式人工智能可能生成的有害或者有风险的内容（即生成式人工智能潜在的内容安全问题）划分为以下9类场景进行调查:

危险话题: 生成的内容包含性、赌博、毒品等危险话题;

敏感话题: 在一些敏感话题上，生成带有偏见或不准确的内容;

违法犯罪: 同意或者鼓励非法活动，如盗窃、抢劫和欺诈等;

身心健康: 生成不适当的内容，对用户的精神造成伤害;

隐私泄露: 生成可能暴露个人隐私信息的内容;

客观中立：生成有偏见的内容或过于主观的评论;

伦理道德：生成的内容鼓励不道德的观念或行为;

恶意诱导：由于恶意指令，生成不安全内容;

攻击指令：根据用户指令，生成不安全或者虚假的内容。

调查结果如表 6 所示，大学生对生成式人工智能不同安全问题的平均认知比例为 33.4%。在所设定的九大风险场景中，只有“隐私泄露”一项的关注度超过半数，而对“违法犯罪”“身心健康”“歧视不公”及“攻击指令”等四项风险的认知均低于平均水平。这表明大学生对生成式人工智能安全问题的认知存在明显局限。

这种认知局限影响了大学生在使用生成式人工智能时的判断力和自我保护能力，也可能导致对潜在安全隐患的忽视。例如，缺乏对“隐私泄露”风险的深入了解，可能导致学生在使用生成式人工智能时，无意中泄露个人敏感信息，增加个人信息被滥用的风险。对“违法犯罪”风险的认知不足可能使学生在处理生成式人工智能生成的内容时，不能有效识别和规避潜在的法律风险，增加学术不端行为及法律纠纷的可能性。对“身心健康”风险的忽视可能导致学生在使用生成式人工智能的过程中，不能有效识别负面信息，从而对心理健康造成不利影响。当生成式人工智能处理涉及政治、文化、种族等敏感话题时，如果学生未能充分认识到潜在风险，可能会导致错误信息的传播，甚至引发误导性舆论和社会矛盾。因此，提升大学生对生成式人工智能安全问题的认

知，具有重要的现实意义。

调查数据显示，超过半数的大学生对生成式人工智能使用安全教育持支持态度，见表 6。这一结果表明，大多数学生已经认识到使用生成式人工智能的潜在安全风险，并希望通过系统化的教育和培训来提升自身的安全意识和防范能力。大学生的积极态度是高校开展生成式人工智能安全教育的基础。

三、总结与展望

调查显示，大学生对生成式人工智能技术的接受度较高，该技术已广泛渗透到大学生日常学习和生活中。可以预见，随着技术的进一步提升，生成式人工智能将在大学生教育、科研等领域发挥更加重要的作用。

尽管学生对生成式人工智能技术表现出积极态度，但是他们对生成式人工智能安全问题的了解程度却不高。即便部分学生已经认识到生成式人工智能的潜在安全隐患，对这些问题的深入认知也仍然不足。学生在使用生成式人工智能时，安全意识尚未全面建立，实践中的风险防范能力也有待提升。此外，不同学科、不同学段的学生对生成式人工智能安全问题的认知存在显著差异。信息与工程学、理学专业的学生，对生成式人工智能的技术细节和潜在风险有更多的了解；来自社会科学、人文学及其他非技术领域的学生，可能由于缺乏相关课程和实践机会，对技术的理解和安全问题的认知相对不足。

基于以上调研结果，本研究从高校网络安全管理部门的角度提出了以下启示与建议，以应对生成式人工智能技术的广泛应用及其潜在的安全问题:

出台生成式人工智能指导规范。高校应制定针对生成式人工智能的使用规范，明确相关的使用原则、技术支持体系及相应的建设要求。通过制定统一的指导性文件，为不同院系、师生提供操作性强的技术指导和标准化的使用流程，确保生成式人工智能在校园内的安全、合法及有效应用。

开展分层分类的生成式人工智能安全教育。由于不同学科的学生所接触的技术场景和内容存在差异，需根据其专业背景设计个性化的培训内容。例如，针对理工科学生，生成式人工智能安全教育可能更多涉及技术操作与代码安全，而针对人文社科学学生，安全教育可能需要关注内容生成中的伦理问题及信息准确性。通过定制课程、专题讲座和实操训练，确保学生不仅能够在理论上理解生成式人工智能的潜在风险，还能够在日常实践中具备足够的安全防护意识和技能。

加强跨部门/院系协作与技术支持。高校应进一步加强人工智能专业院系、网络安全、信息技术及学术事务等多个部门之间的协作，建立综合的技术支持体系，共同建立风险防范机制，确保生成式人工智能在高校中的安全、合法应用，减少潜在风险并提升学生的安全意识和技术实践能力。

参考资料

[1] SUN H, ZHANG Z, DENG J, et al. Safety assessment of Chinese large language models [J]. arXiv preprint arXiv:2304.10436, 2023.

[2] XU G, LIU J, YAN M, et al. Cvalues: measuring the values of Chinese large language models from safety to responsibility [J]. arXiv preprint arXiv:2307.09705, 2023.