

Doi: doi.org/10.70693/rwsk.v1i4.890

DeepSeek 大语言模型在高校落地的伦理风险治理与多维应用价值评估研究

尹立君¹ 宋海军¹ 史雪魁¹ 陈婧¹

(¹ 河北科技工程职业技术大学, 河北 邢台 054035)

摘要: 随着生成式人工智能技术的快速发展, 大语言模型在高校教学、科研与管理等领域的应用日益广泛。作为国产大模型代表, DeepSeek 在落地高校过程中展现出良好的技术适配性与场景应用潜力, 但同时也伴随着数据安全、隐私保护、内容偏见与学术诚信等伦理风险。本文通过系统分析 DeepSeek 的技术架构与高校部署现状, 构建了以技术防护、制度规范与智能监测为核心的伦理风险治理机制, 有效应对多重风险挑战。同时, 构建以教学成效、学生参与和教学创新为核心的多维教学评价模型, 并提出科研创新效能评估体系, 涵盖成果产出、协作效率与创新启发等关键指标。研究成果可为高校在人工智能技术部署、使用与管理中提供科学依据与决策参考, 推动大语言模型技术在高校实现规范、安全与可持续应用。

关键词: DeepSeek; 伦理治理; 教学评估; 科研效能评估

一、引言

1 研究背景

近年来, 以大语言模型 (Large Language Models, LLMs) 为核心的人工智能技术在全球教育领域掀起了系统性变革浪潮。联合国教科文组织 (UNESCO) 在《人工智能与教育: 政策制定者指南》中明确指出, 以生成式人工智能为代表的前沿技术, 通过自适应学习系统、智能辅导工具和知识图谱的深度重构, 正推动着教育范式从“标准化”向“个性化”快速跃迁, 重塑“教”与“学”的内在逻辑与实践模式^[1]。当前, 以 ChatGPT、Gemini 为代表的国际主流大模型已广泛应用于课程设计、作业批改、学习诊断及科研协作等多重教育场景, 显著提升了教育效率与教学效果。

我国高度重视人工智能技术在教育领域的深度融合与创新应用。《新一代人工智能发展规划》(国发〔2017〕35号)明确提出“推动智能教育发展, 构建以学习者为中心的教育环境”^[2]; 教育部发布的《高等学校人工智能创新行动计划》进一步强调, “积极探索人工智能技术支持下的教育教学新模式”, 以促进高校教育实践与技术革新的深度融合^[3]。

DeepSeek 作为国产自研大语言模型的典型代表, 凭借“多模态理解—知识推理—情境化交互”三位一体的创新技术架构, 在高校应用场景中展示出强大的场景适配性与实践价值。教学领域方面, DeepSeek 支持智能备课系统快速生成个性化教案, 能够大幅降低教师在教学准备环节的重复性劳动; DeepSeek 在科研领域通过智能化的文献综述和实验设计建议, 可以有效加快科研周期与成果产出; 在管理领域, DeepSeek 使用 NLP 技术优化行政流程、增强高校内部跨部门沟通与协作效率。已成为推动高校数字化转型的重要工具之一。

2 研究问题与研究意义

作者简介: 尹立君(1980—), 男, 博士, 研究方向为人工智能、大数据分析;

史雪魁(1981—), 男, 硕士, 研究方向为计算机应用、网络技术。

陈婧(1978—), 男, 本科, 研究方向为计算机科学与技术、实验室管理。

通讯作者: 宋海军(1976—), 男, 硕士, 副教授, 研究方向为软件开发;

2.1 研究问题

据统计 2025 年以来已有 50 多所高校进行了 DeepSeek 本地化部署应用,但随之产生的伦理风险逐渐显现,成为高校必须面对的重要问题。一方面,如何准确识别、监控并有效处理 DeepSeek 在教学、科研及管理实践中可能涉及的数据隐私泄露、内容偏见、算法歧视以及价值导向偏差等伦理风险;另一方面,如何构建科学、系统且具有可操作性的大语言模型多维度价值评估框架,以全面衡量其在提升教育质量和推动科研创新的实际成效与潜在影响。

2.2 实践意义

本研究具有显著的现实指导价值。围绕高校在部署 DeepSeek 过程中所面临的伦理治理挑战,将提出一套可行的风险识别与应对机制,为高校管理者提供科学的决策支持。包括:协助高校建立完善的数据安全保护体系,降低信息泄露与隐私侵犯的风险;推动在模型训练与应用过程中引入价值导向审查与偏见识别机制,以减少算法不公所导致的教育不均问题;并进一步为构建高校层面的伦理监督体系提供理论基础与实施建议,助力大语言模型的规范、安全、可持续应用。

另一方面,本文拟构建的“DeepSeek 高校应用多维价值评估模型”,将为高校在模型部署与优化过程中提供量化、系统化的评估工具。模型不仅识别影响模型应用效果的关键因子,实现教学与科研场景的精准匹配与持续改进;为高校管理者在模型选型、资源配置及应用策略制定等方面,提供数据驱动的决策依据,保障师生的合法权益,确保 DeepSeek 技术真正服务于高校教育质量提升与高层次人才培养目标的实现。

二、DeepSeek 在高校落地部署的现状与挑战

2.1 技术架构与实施难点分析

2.1.1 DeepSeek 技术结构解析

DeepSeek 基于 Transformer 深度神经网络结构,结合大规模参数训练与自监督学习方法,具备出色的自然语言理解与生成能力。模型架构采用了改进型混合专家机制 (Mixture of Experts, MoE),通过动态路由策略对计算资源进行智能调度,在保持高性能的同时显著降低了计算消耗,特别适用于资源受限的高校场景。与传统 Transformer 架构相比,DeepSeek 引入稀疏激活机制,使得模型在推理过程中仅需激活约 20% 的参数节点,从而有效控制计算资源消耗并提升处理效率。在专业语言理解方面,模型在学术文本分析任务中能够精准区分高频术语之间的语义边界,例如在医学领域准确识别“细胞凋亡”与“细胞坏死”等术语差异。同时,结合 Prompt 预训练策略与指令微调 (Instruction Tuning) 机制,DeepSeek 在多任务处理与跨学科知识迁移方面展现出优异的泛化能力,为高校提供了智能化、个性化的语言理解与交互支持。

算法原理层面,DeepSeek 通过海量文本语料进行自监督学习,综合采用掩码语言建模 (Masked Language Modeling, MLM) 和因果语言建模 (Causal Language Modeling, CLM) 方法,协同实现语言理解和文本生成的双重目标。同时,DeepSeek 引入了基于人类反馈的强化学习方法 (Reinforcement Learning from Human Feedback, RLHF),通过人类偏好反馈不断优化模型表现,使其在复杂指令理解、人机交互等任务中展现更优异的性能。

数据处理策略层面,DeepSeek 具有高效的数据预处理和并行计算方案,支持超大规模语料库的并行训练和分布式学习,能够快速响应高校多元化、实时化的应用需求。此外,DeepSeek 还强化了少样本学习 (Few-shot Learning) 和零样本学习 (Zero-shot Learning) 机制,帮助模型快速泛化到各类教育教学和科研任务中,可以有效降低高校部署应用过程中的训练与调优成本。

2.1.2 与高校现有技术系统的兼容性挑战

尽管 DeepSeek 在技术层面具有显著的先进性与灵活性,但在实际部署到高校现有技术生态中时,也面临诸多兼容性挑战。一是高校现有教务系统普遍使用基于关系型数据库 (如 Oracle、MySQL) 的数据管理模式,而 DeepSeek 则依赖于图数据库 (如 Neo4j) 实现知识关联与推理,两类数据库在数据模式转换过程中容易产生结构化数据损失,影响知识图谱的完整性;二是高校计算资源主要基于传统 CPU 集群架构,而 DeepSeek 所需的 GPU 或 TPU 异构计算资源在配置上存在明显冲突,增加了运行成本与能耗;最后,高校主流教学管理软件 (如 Blackboard、Moodle) 所采用的 RESTful 接口与 DeepSeek 基于 gRPC 的接口协议之间存在技术差距,导致系统集成困难,降低了模型应用的实时响应速度与用户体验。

2.2 实施过程中的技术难点应对策略

针对上述兼容性挑战,在实施 DeepSeek 过程中,高校可采取以下应对策略:一是构建统一的数据接口标准平台,通过制定数据交互规范和部署基于中间件的数据转换引擎(如 Apache NiFi),实现多源数据的实时高效转换,降低异构数据之间的结构化损失,有效解决数据孤岛问题。二是采用分层架构与容器化部署技术,以微服务架构与容器技术(如 Kubernetes)实现 DeepSeek 服务与现有高校系统的融合,既确保新旧系统的并行运行稳定性,又高效实现 CPU 和 GPU 等异构算力资源的动态调度与优化。三是建立严格的数据治理与安全合规机制,通过明确数据访问权限、强化数据加密与脱敏措施以及建立数据访问审计体系,保障数据安全和隐私合规,降低伦理风险。

三、伦理风险治理机制构建

3.1 数据安全与隐私风险治理机制

3.1.1 数据安全的技术保障路径

为有效规避 DeepSeek 在部署与应用过程中面临的数据安全隐患,尤其是数据泄露、恶意篡改及非授权访问等风险,高校构建多维度的技术保障体系,综合运用数据脱敏、访问控制与加密保护等手段,全面提升数据安全防护水平。

数据脱敏技术作为保护敏感信息的重要手段,通过模糊化、匿名化、混淆化及泛化处理等方式,有效阻断对师生个人敏感信息的直接识别与非法利用。以清华大学智慧校园系统为例,广泛部署的数据脱敏机制在防范师生信息泄露方面取得显著成效,为其它高校数据治理提供了可借鉴范式。

在权限管理方面,遵循“最小权限”原则,构建严密的访问控制策略,确保数据访问权责清晰、机制完备。基于采用基于角色的访问控制(RBAC)或基于属性的访问控制(ABAC)等技术方法,实现对数据访问行为的精准管控与实时追踪。

此外,在数据存储与传输环节,高校需采取强加密策略保障数据链路安全。通过部署对称加密算法(如 AES)与安全传输协议(如 TLS、SSL),可有效抵御数据在传输与存储过程中的篡改与窃取风险,确保数据完整性与保密性。

3.1.2 隐私保护制度的构建与规范

完备的制度保障体系是实现高校师生数据隐私保护的关键所在。高校在应用 DeepSeek 等大语言模型的过程中,严格按照《中华人民共和国个人信息保护法》《数据安全法》等相关法律法规,明晰个人数据在采集、存储、处理、使用、共享及删除等生命周期环节的具体操作规范,构建覆盖全过程的隐私治理框架。明确各方责任主体,建立数据隐私责任人制度,厘清模型提供方、高校信息化部门与具体使用者在数据治理中的职责边界。同时,推动形成“动态透明”式告知机制,通过公开数据用途、使用范围、技术方式及潜在风险,确保师生在充分知情基础上参与数据授权与交互。

3.2 学术诚信和内容真实性风险防范体系

3.2.1 学术诚信智能监测机制的建设

习近平总书记在 2024 年全国科技大会等场合明确提出要加强科研诚信和作风学风建设,“科研诚信是科技创新的基石”。高校师生在使用 DeepSeek 过程中可能引发学术不端行为,如抄袭、代写、剽窃等现象,亟需构建以技术为核心支撑的学术诚信监测体系。引入文本相似度检测、引文网络分析与作者风格识别等智能化工具,对师生提交的学术作品与现有文献资源进行多维比对与动态监控,实现学术不端行为的实时预警与风险管控。例如,华北电力大学、湖北大学及福州大学等高校已先行部署了基于生成式人工智能(AIGC)识别机制的学术检测系统,使用 AI 辅助相似度分析与语体风格比对,降低了学术成果造假、剽窃现象的发生率。

3.2.2 学术规范治理与违规行为处置

3.2.2 智能生成内容的质量控制与学术治理体系融合架构

为应对大语言模型在高等教育应用中引发的学术风险,需建立贯穿内容生成全周期的“质量管控-规范治理”协同机制。在质量管控维度,实施三级递进式审核体系:基础层通过自然语言处理技术筛查逻辑冲突与事实性错误,中间层依托跨学科专家团队优化内容结构与学术表达,终审层对接权威知识库实现多源交叉核验,从而系统性保障生成内容的准确性与规范性。同步构建智能学术治理框架,将传统学术伦理准则迭

代为涵盖 AI 创作特征的《生成式内容应用规范》，如《复旦大学关于在本科毕业论文（设计）中使用 AI 工具的规定（试行）》，明确界定 AI 辅助抄袭、虚构数据等新型学术不端行为的界定标准。同时建立“智能监测 - 人工核查 - 分级惩戒”的处置链条，实现违规行为可追溯、可定责。多所高校已将学术诚信专题纳入新生入学教育、研究生开题培训及教师继续教育课程，有效提升了全校范围内的学术自律水平与制度执行力。

四、DeepSeek 高校应用的多维价值评估模型构建

4.1 教学质量提升的评价模型设计

4.1.1 指标体系的构建依据

构建教学质量评估体系时，综合借鉴了 Kirkpatrick 四层次评估模型、Community of Inquiry (CoI) 框架以及生成式人工智能在教育领域的最新应用成果，从多维度全面评估 DeepSeek 在教学实践中的实际效能^[4]。最终形成的指标体系如表 1 所示，涵盖教学成效、学生参与度与教学方法创新三大维度。

在“教学成效”维度，核心评价内容包括学生对知识的掌握情况、技能水平的提升幅度以及课程教学目标的完成程度。“学生参与度”指标聚焦学生在课堂教学中的活跃程度与主动性表现，主要通过课堂互动频率、学习动机强度以及自主探究行为等要素进行度量。“教学方法创新”评估 DeepSeek 技术在推动教学模式改革中的具体作用，涉及教学手段优化、教学工具革新以及教师教育理念更新等方面。

表 1 教学质量评估体系

一级指标	权重 (%)	二级指标	权重 (%)
教学效果	50	学生成绩提升情况	25
		学生知识掌握程度	15
		教学目标达成率	10
学生参与度	30	学生课堂互动情况	15
		学生学习积极性与主动性	15
教学方法创新	20	教学手段创新性	10
		教学模式优化情况	10

4.1.2 确定指标权重的方法

为了保证评估体系中各项指标权重的分配科学、客观且具可操作性，研究结合了德尔菲法 (Delphi Method) 和层次分析法 (AHP) 来确定指标权重^[5]。

第一步，邀请具有代表性的教育学专家及高校一线教师开展多轮德尔菲法咨询，通过专家意见汇总与匿名反馈，初步确定各指标的相对重要性。第二步，基于德尔菲法所得结果，构建层次结构模型，并运用 AHP 方法构建判断矩阵，执行一致性检验，从而完成对各项指标权重的量化计算。

具体显示结果：“教学成效”维度权重最高，占比 50%，强调教学产出的核心地位，契合余“以学习成效为导向”的教育技术应用原则^[6]；“学生参与度”权重为 30%，强调教学过程中的动态反馈与学习过程质量，与 Fredricks 所提出的行为、情感与认知三维度参与理论相契合^[7]；“教学方法创新”则占据 20% 的权重，用以衡量人工智能技术推动教学方式变革的深度与广度，符合《教育信息化 2.0 行动计划》中关于“技术引领教学创新”的战略要求。

4.1.3 评价模型构建与应用

基于上述评价指标与权重，研究构建了 DeepSeek 教学质量提升的综合评价模型。分为一级指标和二级指标两个层次，通过权重计算得到最终的综合评分，以评估 DeepSeek 技术在教学实践中的实际表现。综合评分具体计算公式如下：

其数学表达式为： $Q=0.5E+0.3P+0.2I$

其中，Q 为教学质量指数，E、P、I 分别为教学效果、学生参与度、教学方法创新的标准化得分 (0-100 分)

在模型的实际应用过程中，具体通过课堂观察记录、问卷调查、成绩分析、访谈反馈等多种途径，对每个评价指标逐一进行详细评分，再依据上述公式进行权重计算，最终获得代表 DeepSeek 技术综合教学质

量提升效果的总评分。

4.2 科研创新支持效能评估指标体系

4.2.1 指标体系框架构建

科研创新支持效能评估指标体系需综合考虑 DeepSeek 在促进科研成果产出、提升科研合作效率、启发科研思维等方面的作用。包括科研成果数量与质量（如论文发表数量及引用率、项目申报成功率）、科研合作频率及效率、跨学科合作情况以及科研人员对新方法、新思路的接受与应用程度。基于科技部《国家创新驱动发展战略纲要》提出的“创新链-产业链-价值链”协同理论^[8]，结合生成式 AI 的科研赋能特性，构建科研创新支持效能评估指标体系（见表 2）。

表 2 科研创新支持效能评估指标体系

一级指标（权重）	二级指标	权重	数据来源	测量工具
科研成果产出（40%）	学术论文质量	15%	WoS/Scopus 数据库	影响因子加权算法
	专利转化效率	10%	国家知识产权局公示数据	收益-成本比模型
	科研项目层级	15%	科研管理系统	NSFC 项目分级标准
科研合作效率（35%）	跨学科合作指数	12%	合作网络图谱	模块化指数（Q 值）计算
	国际合作频次	10%	国际合作项目清单	频次-影响力复合指标
	人机协作深度	13%	系统操作日志	行为序列挖掘技术
	创新路径突破性	10%	技术预见报告	颠覆性创新指数（TDI）
科研思维启发（25%）	研究范式转型度	8%	Altmetrics 数据	社交媒体传播力分析
	科研人员认知跃迁	7%	认知行为实验数据	脑电信号复杂度分析

4.2.2 指标数据收集与分析方法

数据收集与分析方法应结合多种手段确保数据可靠性与有效性，可采用“多源异构数据融合+混合研究方法”。具体包括文献计量分析（如通过 Web of Science、CNKI 等数据库进行科研成果文献分析）、科研人员问卷调查（评估科研人员的主观感受及满意度）、科研项目跟踪（动态监测项目进展、合作交流和技术应用情况）。

4.2.3 评估结果应用与反馈

科研创新支持效能评估结果积极应用于科研管理决策和科研人员激励机制之中。通过评估结果，更准确地识别 DeepSeek 在科研领域的应用优势与不足，制定更加精准的激励政策和资源配置方案，建立“评估-决策-优化”闭环管理系统，将效能评估深度嵌入科研管理全流程。另外，建立有效的反馈机制，根据评估反馈结果，调整优化 DeepSeek 技术在科研实践中的应用策略，推动高校科研能力持续提升。

五、研究结论

本研究聚焦 DeepSeek 大语言模型在高校落地应用过程中的伦理风险治理与多维应用价值评估，得出了以下主要结论并形成了若干创新性成果：DeepSeek 作为新一代国产大语言模型，凭借多模态理解、知识推理和情境化交互三大核心优势，具备显著的技术先进性与高校应用适配性，其在教学、科研与管理等场景中展现出高效能与强泛化能力。针对高校在部署过程中面临的数据安全、隐私保护与学术诚信风险，本文构建了涵盖数据脱敏、权限控制与加密保护在内的技术体系，同时辅以《个人信息保护法》等法规指引下的隐私治理制度与学术规范监测机制，实现了从技术到制度的综合治理路径。本研究系统性构建了高校教学质量提升评价模型与科研创新支持效能评估框架，前者基于 Kirkpatrick 模型与 CoI 理论，融合教学成效、学生参与度与教学方法创新三大维度，后者围绕科研产出、合作效率与创新思维三要素。在理论层面，本研究填补了国内在高校生成式人工智能伦理治理与价值评估领域的研究空白；在实践层面，为高校部署与管理大语言模型技术提供了可操作、可推广的模型与策略，具有广泛的现实应用价值。

参考文献:

- [1] UNESCO. AI and Education: Guidance for Policy-makers[R]. Paris: UNESCO Publishing, 2021.
- [2] 国务院. 新一代人工智能发展规划[Z]. 国发〔2017〕35号, 2017.
- [3] 教育部. 高等学校人工智能创新行动计划[Z]. 教技〔2018〕5号, 2018.
- [4] Lee J, Toutanova K. Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018, 3(8).
- [3] 张小艳, 段正宇. 基于句级别 GAN 的跨语言零资源命名实体识别模型[J]. 计算机应用, 2023, 43(08): 2406-2411.
- [4] 龚苏宁, 陈荣华, 吴耀华. 基于 Kirkpatrick 模型的高校艺术设计专业 SPOC 混合式教学效果评估模型构建研究[J]. 美与时代 (上旬刊), 2022(12): 142-147.
- [5] 傅盈盈. 人工智能赋能高校思想政治教育的技术逻辑、价值意蕴和实践路径[J]. 湖北第二师范学院学报, 2024, 41(07): 14-19.
- [6] 曹建磊, 黄迎辉, 吕部. 基于学习成效的地方应用型高校教学质量保障体系建设研究[J]. 黑龙江工业学院学报(综合版), 2024, 24(05): 1-6.
- [7] Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research, 74, 59-109.
- [8] 中共中央, 国务院. 国家创新驱动发展战略纲要 [EB/OL]. (2016-05-19). http://www.gov.cn/zhengce/2016-05/19/content_5074812.htm.

A Study on the Ethical Risk Governance and Multidimensional Application

Value Assessment of the DeepSeek Large Language Model in Higher Education

Yin Lijun, Song Haijun, Shi Xuekui, Chen Jing

Hebei Vocational University of Technology and Engineering, Xingtai, Hebei 054035, China

Abstract: With the rapid advancement of generative artificial intelligence, large language models (LLMs) have been increasingly applied in higher education, including teaching, research, and administrative management. As a representative domestic LLM, DeepSeek has demonstrated strong technical adaptability and application potential in various university contexts. However, its deployment also raises ethical concerns related to data security, privacy protection, content bias, and academic integrity. This paper systematically analyzes the technical architecture of DeepSeek and its current deployment in higher education institutions. Based on the findings, it proposes an ethical risk governance framework centered on technical safeguards, institutional norms, and intelligent monitoring to effectively address multiple layers of risk. Furthermore, the study develops a multidimensional teaching evaluation model focusing on teaching effectiveness, student engagement, and instructional innovation. It also establishes a research innovation performance evaluation framework, incorporating key indicators such as research output, collaboration efficiency, and innovation stimulation. The research outcomes aim to provide a scientific basis and decision-making reference for the deployment, utilization, and governance of AI technologies in universities, thereby promoting the standardized, secure, and sustainable application of large language models in higher education.

Keywords: DeepSeek; Ethical Governance; Teaching Evaluation; Research Performance ; Assessment