Creative Education Studies 创新教育研究, 2025, 13(5), 244-250
Published Online May 2025 in Hans. https://www.hanspub.org/journal/ces
https://doi.org/10.12677/ces.2025.135332

Hans汉斯

大语言模型在教育教学过程中的伦理道德挑战与应对策略

杨文彬

西安邮电大学理学院，陕西 西安

收稿日期：2025年3月21日；录用日期：2025年5月6日；发布日期：2025年5月15日

摘　要

大语言模型在教育领域的广泛应用推动了教学方式的革新，但也带来了许多道德与伦理挑战。本文分析了大语言模型在教育教学中存在的数据隐私、算法偏见和学术不诚信等问题，探讨了建立伦理框架与监管机制的重要性，并提出了一系列改进措施以确保其安全、合理地应用。通过研究这些问题与对策，本文为推动教育技术的合规性和伦理性提供了理论依据和实践指导。

关键词

大语言模型，教育教学，道德挑战，伦理问题，改进对策

Ethical and Moral Challenges and Response Strategies of Large Language Models in the Educational Teaching Process

Wenbin Yang

School of Science, Xi'an University of Posts and Telecommunications, Xi'an Shaanxi

Received: Mar. 21st, 2025; accepted: May 6th, 2025; published: May 15th, 2025

Abstract

The widespread application of large language models in the field of education has driven innovation in teaching methods, but it has also brought about numerous ethical and moral challenges. This paper analyzes issues such as data privacy, algorithmic bias, and academic dishonesty present in the use of large language models in education. It explores the importance of establishing ethical frameworks and regulatory mechanisms and proposes a series of improvements to ensure their safe and

文章引用: 杨文彬. 大语言模型在教育教学过程中的伦理道德挑战与应对策略[J]. 创新教育研究, 2025, 13(5): 244-250.
DOI: 10.12677/ces.2025.135332

reasonable application. By studying these issues and countermeasures, this paper provides theoretical foundations and practical guidance for promoting the compliance and ethics of educational technology.

Keywords

Large Language Models, Educational Teaching, Ethical Challenges, Moral Issues, Improvement Measures

1. 引言

随着人工智能(AI)技术的迅猛发展，尤其是大语言模型(如 DeepSeek、OpenAI 的 ChatGPT 和 Google 的 BERT 等)，在教育领域的应用已成为一种不可忽视的趋势[1]-[3]。大语言模型能够通过自然语言处理技术，生成高质量的文本，解答学术问题，并辅助教师进行教学决策，极大地提高了教育效率与质量。例如，这些模型能够为学生提供个性化学习建议，帮助教师更好地了解学生的学习进度和需求，从而优化教学方案[4] [5]。

然而，语言模型在教育中提供了众多积极应用的同时，其带来的道德与伦理问题也愈加突出。这些问题涉及数据隐私、算法偏见、学术不诚信等方面，若处理不当，可能对教育公平、学生权益以及学术道德带来潜在威胁[6]。

近年来，学术界和技术界对这些问题展开了广泛讨论。例如，数据隐私问题指的是大语言模型在训练过程中使用大量学生数据，如何确保这些数据的安全性和匿名性是亟待解决的课题[3]。此外，由于训练数据和算法的局限性，算法可能会产生偏见，影响到教育中的公平性[5]。再者，学术不诚信问题的出现，也使语言模型的应用面临挑战，学生利用 AI 自动生成的内容完成作业或论文，这可能削弱教育体系的诚信原则[7]。

因此，本文将从以下几个方面深入探讨大语言模型在教育中面临的道德问题，并提出相应的解决对策：首先，分析当前大语言模型在教育中所面临的主要伦理挑战；其次，探讨如何建立有效的伦理框架与监管机制，确保这些技术的安全和合理应用；最后，提出改进措施，促进大语言模型在教育中的合规性与伦理性，保障学生的权益和教育公平性。

2. 大语言模型在教育中的道德与伦理问题

2.1. 数据隐私问题

大语言模型的训练需要海量的、具有多样性的文本数据，而在教育领域，这些数据往往涉及学生、教师的个人信息、学习行为、成绩等敏感内容。如果数据存储和使用没有得到严格管理，可能会导致隐私泄露和滥用[8] [9]。尤其是当教育平台和 AI 公司没有采取足够的隐私保护措施时，学生的学习过程和个人信息可能被不当采集、存储或出售。

AI 技术在教育领域中的数据隐私问题不仅仅涉及数据泄露的风险，还包括“数据追踪”的问题[10]。许多教育平台在收集学生数据时可能未能完全告知学生和家长，这在某些情况下违反了数据保护法(如 GDPR)规定。教育平台和 AI 系统收集的数据如果未经过加密或去标识化处理，容易被用于不当分析，甚

至可能被用于商业化推广或行为预测[9]。许多教育 AI 平台未能妥善处理学生的数据，导致数据泄露事件频发，并引发家长和社会对教育隐私保护的广泛担忧。因此，如何确保大语言模型在教育中的数据隐私保护成为一个亟待研究和规范的关键问题。

2.2. 算法偏见问题

大语言模型通常依赖大量的历史数据进行训练，而这些数据可能携带历史偏见，导致算法本身也可能展现出偏见[11] [12]。例如，若训练数据包含性别、种族或地域等方面的偏见信息，模型可能在生成内容时加剧教育资源的不平等分配，甚至在某些情况下对少数群体、特定性别或其他弱势群体产生歧视性输出。这不仅影响学生的学习体验，也可能加剧教育系统本身的公平性问题。

算法偏见问题在教育 AI 系统中尤为严重。教育领域中的大语言模型如果依赖于不平衡的训练数据，可能会在评分、推荐和个性化学习路径设计上加剧社会不公[11] [12]。例如，某些性别或族裔背景的学生可能更容易遭遇负面的评价，或者教育资源的推荐存在不公平分配。这种算法偏见不仅在教育内容的生成上产生影响，在学生成绩预测等领域也可能存在严重的不公。偏见不仅仅体现在模型生成内容的表面，深层次的偏见问题可能影响教师的决策，导致在教务管理、招生选拔等环节的不公。因此，在开发和使用大语言模型时，必须采取措施纠正算法中的偏见，确保模型的公平性与透明度。

2.3. 学术不诚信问题

大语言模型的另一个重要道德和伦理问题是学术不诚信。随着 AI 技术的快速发展，学生使用大语言模型生成作业、论文甚至考试答案的现象日益增加。这种行为不仅挑战了学术诚信的基本原则，还可能改变传统教育体系中作业评价和考试评估的标准[13]。

大语言模型生成的内容在某些情况下已经可以骗过传统的学术检测工具，这使得学术不诚信问题更加复杂[14] [15]。学生通过利用模型自动生成的内容，可以在短时间内获得高质量的作业或论文，而这些内容往往缺乏独立思考和原创性。AI 生成的内容虽然看似符合学术要求，但由于缺乏对复杂问题的深入理解，最终可能导致学生的思维能力和分析能力的退化。

虽然大多数教育机构采取了防范措施，但 AI 生成内容的检测能力仍有不足，甚至部分学生使用 AI 工具来“优化”其作业内容，从而绕过学术检测机制[15]。这种行为不仅危及学生的学术成长，也影响教育机构对学生能力的真实评价，进而影响教育质量的公正性。因此，学术不诚信问题已经成为教育系统中需要重视的伦理难题，呼吁更为严格的监管和教育政策，以遏制 AI 滥用的趋势。

综上所述，大语言模型在教育中的应用虽然带来了诸多便利，但也伴随着数据隐私、算法偏见和学术不诚信等一系列伦理道德风险。这些问题不仅威胁到教育系统的公正性与可信度，也对教育参与者的权利和成长造成潜在影响。因此，为了在保障教育质量的同时合理利用 AI 技术，需要及时构建系统性的伦理框架与有效的监管机制，以规范大语言模型在教育场景中的发展路径与应用边界。

3. 建立伦理框架与监管机制

3.1. 制定道德原则

在大语言模型的教育应用中，制定明确的道德原则是确保其合规和伦理应用的基石。这些道德原则应以公平、透明和责任为核心，特别是在数据隐私、算法偏见以及学术不诚信等方面进行规范。据 NGUYEN 等人的研究[16]，公平性原则要求确保大语言模型在教育过程中对所有学生一视同仁，避免任何形式的歧视性输出，特别是性别、种族或其他社会经济背景上的偏见。透明性原则要求教育者和学生能够了解模型的决策过程，确保模型的工作原理和数据来源能够公开，并接受公众监督。责任原则强调

AI 技术开发者和教育机构应对模型的使用承担责任，包括教育环境中的潜在误用或不当应用。

例如，依据公平性原则，教育机构应禁止通过大语言模型收集和使用敏感数据，如学生的家庭背景、健康信息等，除非明确获得学生或家长的同意。同时，禁止利用大语言模型进行不当教学操作，如生成可能误导学生的答案或在评估过程中利用 AI 系统对学生进行不公正的评分[17]。研究还指出，教育系统应建立严格的道德准则，确保大语言模型的使用不对学生的心理健康、社交互动或价值观形成负面影响[13]。

3.2. 建立多层次监管体系

有效的监管机制是确保大语言模型安全、合规应用的关键。建立多层次的监管体系，包括校内监管和社会监管，可以从不同层面保障其伦理性。首先，学校应设立技术伦理委员会或数据伦理委员会，专门负责审查 AI 系统在教学中的应用，确保这些应用符合道德规范，并避免潜在的风险。伦理委员会可以定期评估大语言模型的使用效果，确保其在帮助学生学习的同时不会引发学术不诚信、隐私泄露等问题[18]。

其次，国家层面应出台相应的法律法规，明确大语言模型在教育领域的应用边界。这些法规应详细规定教育机构在使用 AI 时的责任和义务，特别是如何处理学生数据、如何预防算法偏见、如何确保学术诚信等。最新研究表明，许多国家已经开始推动“AI 教育法”的立法工作，以规范 AI 在教育领域中的使用，保护学生的权益，并促进教育公平[19]。例如，欧盟在 2023 年通过了《人工智能法案》(AI Act)，要求所有 AI 应用必须进行风险评估，并在高风险应用领域(如教育)进行更加严格的监管。该法案强调，AI 系统的透明性和可解释性是法规中的核心内容，同时要求教育机构对其 AI 系统的使用情况进行定期审查和报告[17]。

3.3. 增强技术透明性

技术透明性是保障大语言模型在教育领域合规应用的重要措施之一。推广模型可解释性研究，要求开发者披露模型训练数据和算法偏差来源，可以有效减少“黑箱效应”带来的伦理隐患。据 KASTANIA 的研究[20]，AI 模型的“黑箱效应”指的是由于模型复杂性过高，导致开发者和用户难以理解模型的具体决策过程，从而无法识别其潜在的偏见和错误。

近年来，许多学者和开发者已经开始致力于提高大语言模型的可解释性，提出了多种方法来揭示模型决策背后的原因，如通过局部可解释性模型(LIME)和 SHAP 值来解释模型的具体决策[20]。此外，要求开发者披露模型训练数据的来源和使用情况，以及可能存在的算法偏见，是提升透明性的重要举措[19]。2023 年，学术界和业界对这一问题的关注有所增加，许多教育技术公司已开始在其产品中加入可解释性功能，以帮助教师和教育管理者更好地理解 AI 模型的决策逻辑。

增强透明性的另一个重要方面是，要求开发者公开 AI 模型训练时使用的数据集，尤其是涉及学生数据的部分。只有当数据来源、训练方法和使用场景完全透明时，用户(包括学生、家长和教师)才能更好地信任这些系统，并在必要时对其进行监督和纠正。

为实现大语言模型在教育中的可持续发展，不仅要依靠制度和技术层面的规范，还应通过教育者的主动参与、技术手段的合理使用以及数据保护意识的强化来保障其正向价值的实现。下一节将从应用实践出发，提出具体的改进措施与建议，以应对当前面临的伦理挑战，推动技术在教育领域中的健康发展。

4. 改进措施与应用建议

随着大语言模型在教育领域的日益普及，确保其道德与伦理性应用不仅需要建立健全的监管机制，还需要从多个角度实施切实可行的改进措施。以下结合最新的研究文献，提出具体的改进措施与应用建

议，帮助教育领域更好地应对与解决大语言模型带来的道德与伦理挑战。

4.1. 加强教育者的技术素养

教育者的技术素养是确保大语言模型在教育中合理应用的关键。教师和教育管理者需充分理解大语言模型的局限性与潜在风险，从而避免其不当使用。

教师的 AI 素养不仅要包括如何使用这些技术工具，还要理解其可能带来的隐私泄露、算法偏见及学术不诚信等问题[21]。因此，建议定期组织教师和教育管理者参加相关的培训和讲座，尤其是针对 AI 伦理和隐私保护的专题培训。这些培训应包括以下几个方面：第一，教授教育者如何合理使用大语言模型辅助教学，如为学生提供即时反馈、个性化学习建议等；第二，帮助教育者识别 AI 在教学中的潜在偏见及局限性，如何避免过度依赖 AI 技术以保证教育质量；第三，强化教师的伦理意识，确保他们理解在教学过程中如何保护学生的数据隐私与安全。

通过 AI 素养培训，教师不仅提升了技术使用的能力，也增强了他们的伦理判断力，能够在面对 AI 系统的应用时作出更合适的决策[22]。

4.2. 引入防范学术不诚信的技术

随着大语言模型在教育中的广泛应用，学生可能通过 AI 生成作业、论文或答案，从而降低学术诚信的标准。为此，开发针对大语言模型生成内容的检测工具是必不可少的。

学者们应开发多种基于机器学习的检测工具，用以识别 AI 生成的文本。例如，通过分析文本中的句式结构、语法模式以及重复性内容，能够有效区分由 AI 生成和人类创作的作品[23]。教育平台应积极引入此类检测工具，在学生提交作业或论文时，对其进行原创性验证。通过使用这类技术，教师可以迅速识别是否有学生借助 AI 进行作业代写，并引导学生正确使用 AI 工具，促进学术诚信。

此外，学校应教育学生如何负责任地使用 AI 工具，将其作为学习辅助，而非代替学习的工具。通过引导学生合理使用 AI，培养他们的学术诚信意识，进一步减少不当使用 AI 生成内容的情况[24]。

4.3. 建立学生数据保护机制

学生数据的保护是确保大语言模型合规应用的另一关键环节。AI 模型需要海量数据进行训练，而教育数据往往涉及学生的个人信息，如姓名、成绩、健康状况等。若这些数据被滥用或泄露，可能对学生造成严重的隐私侵犯。

为了保护学生的个人数据，教育机构可以采用数据加密、匿名化处理等技术。通过对敏感数据进行加密处理，确保数据在存储和传输过程中的安全性；采用匿名化处理时，尽量避免收集能够直接识别学生身份的信息。数据加密和匿名化处理是当前大多数教育平台数据保护的基本手段[25]。

此外，学校应定期引入第三方机构进行数据安全审查，确保教育平台符合数据保护法规。近年来，部分教育平台已经开始委托第三方进行数据安全评估，并定期发布安全报告，以增强学生、家长和公众对数据安全的信任[26]。

4.4. 强调技术应用的教育公平性

教育技术的应用不仅要追求高效，还要注重教育公平性。为了避免大语言模型在教育中可能带来的偏见问题，设计和应用过程中应特别关注数据的多样性和包容性，确保模型的输出不会加剧已有的社会不公。

教育技术中的算法偏见往往源于训练数据的单一性或不均衡性。例如，若训练数据主要来自城市地区，模型在处理农村地区学生的问题时，可能无法准确理解其背景，导致教育资源分配的不公平。因此，

在设计大语言模型时，应注重数据的多样性与包容性，确保不同地区、不同背景的学生在使用过程中能够获得公平的教育资源。

此外，在教育资源分配中，尤其是技术相对落后的地区，应借助人工智能技术，确保所有学生无论起点如何，都能够享有平等的进步机会，并实现超越[27]。例如，一些教育技术公司已开始与农村学校合作，向这些学校提供基于 AI 的教育资源和培训，帮助学生提高学习效果。政府和社会也应出台政策，鼓励对贫困地区和教育资源薄弱地区进行技术扶持，确保所有学生都能平等享受教育的成果。

5. 结论

大语言模型在教育领域的广泛应用，为教学方式、学习路径和教育资源的优化带来了全新可能，正加速推动教育智能化转型。然而，伴随技术深入发展，数据隐私泄露、算法偏见、学术诚信缺失等伦理问题也日益显现，需要及时建立系统的治理机制加以应对。为此，需从透明性、公平性和责任性等维度构建伦理框架，通过政策制定、技术规范与教育实践的协同，提升人工智能在教育场景中的可信性与可控性。

面向未来，大语言模型赋能教育的前提是实现技术进步与伦理治理的良性互动。应推动跨学科合作，强化法律、技术与社会因素的联动响应，提升数据治理水平与算法公正性，确保教育公平与可持续发展。唯有在多方合力下，才能释放人工智能在教育变革中的真正价值，实现高质量发展目标。

基金项目

2023 年陕西省教育教学改革研究项目“创新人才培养的大学数学课程教学改革与探索”(23BY096)、2024 年西安邮电大学教学改革研究专项项目“《数学物理方法》课程中思政教育改革的策略与实践研究”(JGSZB202419)。

参考文献

[1] Dreyer, J. (2025) China Made Waves with Deepseek, but Its Real Ambition Is AI-Driven Industrial Innovation. *Nature*, **638**, 609-611. https://doi.org/10.1038/d41586-025-00460-1

[2] Uang, X., Zou, D., Cheng, G., *et al*. (2023) Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology & Society*, **26**, 112-131.

[3] 刘明, 吴忠明, 廖剑, 等. 大语言模型的教育应用: 原理、现状与挑战——从轻量级 BERT 到对话式 ChatGPT [J]. 现代教育技术, 2023, 33(8): 19-28.

[4] Saaida, M.B.E. (2023) AI-Driven Transformations in Higher Education: Opportunities and Challenges. *International Journal of Educational Research and Studies*, **5**, 29-36.

[5] 吴永和, 姜元昊, 陈圆圆, 等. 大语言模型支持的多智能体: 技术路径、教育应用与未来展望[J]. 开放教育研究, 2024, 30(5): 63-75.

[6] 王峰. 人工智能需要“灵魂”吗——由大语言模型引发的可能性及质疑[J]. 上海师范大学学报(哲学社会科学版), 2023, 52(2): 5-13.

[7] Fu, Y. and Weng, Z. (2024) Navigating the Ethical Terrain of AI in Education: A Systematic Review on Framing Responsible Human-Centered AI Practices. *Computers and Education*: *Artificial Intelligence*, **7**, Article ID: 100306. https://doi.org/10.1016/j.caeai.2024.100306

[8] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., *et al*. (2023) Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *British Journal of Educational Technology*, **55**, 90-112. https://doi.org/10.1111/bjet.13370

[9] Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., *et al*. (2023) Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models. *Meta-Radiology*, **1**, Article ID: 100017. https://doi.org/10.1016/j.metrad.2023.100017

[10] Horzyk, A. (2023) Data Protection and Privacy: Risks and Solutions in the Contentious Era of AI-Driven Ad Tech. In: Luo, B., Cheng, L., Wu, Z.G., Li, H. and Li, C., Eds., *Neural Information Processing*, Springer, 352-363.

https://doi.org/10.1007/978-981-99-8181-6_27

[11] Idowu, J.A., Koshiyama, A.S. and Treleaven, P. (2024) Investigating Algorithmic Bias in Student Progress Monitoring. *Computers and Education*: *Artificial Intelligence*, **7**, Article ID: 100267. https://doi.org/10.1016/j.caeai.2024.100267

[12] Chinta, S.V., Wang, Z., Yin, Z., *et al*. (2024) FairAIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications. arXiv: 2407.18745.

[13] Fowler, D.S. (2023) AI in Higher Education: Academic Integrity, Harmony of Insights, and Recommendations. *Journal of Ethics in Higher Education*, No. 3, 127-143. https://doi.org/10.26034/fr.jehe.2023.4657

[14] Xie, Y., Wu, S. and Chakravarty, S. (2023) AI Meets AI: Artificial Intelligence and Academic Integrity—A Survey on Mitigating AI-Assisted Cheating in Computing Education. *The* 24*th Annual Conference on Information Technology Education*, Marietta, 11-14 October 2023, 79-83. https://doi.org/10.1145/3585059.3611449

[15] Park, H.E. (2024) The Double-edged Sword of Generative Artificial Intelligence in Digitalization: An Affordances and Constraints Perspective. *Psychology & Marketing*, **41**, 2924-2941. https://doi.org/10.1002/mar.22094

[16] Nguyen, A., Ngo, H.N., Hong, Y., Dang, B. and Nguyen, B.T. (2022) Ethical Principles for Artificial Intelligence in Education. *Education and Information Technologies*, **28**, 4221-4241. https://doi.org/10.1007/s10639-022-11316-w

[17] European Commission (2021) Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM (2021) 206 Final.

[18] Schiff, D. (2021) Education for AI, Not AI for Education: The Role of Education and Ethics in National AI Policy Strategies. *International Journal of Artificial Intelligence in Education*, **32**, 527-563. https://doi.org/10.1007/s40593-021-00270-2

[19] Radu, R. (2021) Steering the Governance of Artificial Intelligence: National Strategies in Perspective. *Policy and Society*, **40**, 178-193. https://doi.org/10.1080/14494035.2021.1929728

[20] Kastania, N.P. (2024) Building Trust in AI Education: Addressing Transparency and Ensuring Trustworthiness. In: Kourkoulou, D., Tzirides, A.O., Cope, B. and Kalantzis, M., Eds., *Trust and Inclusion in AI-Mediated Education*, Springer, 73-90. https://doi.org/10.1007/978-3-031-64487-0_4

[21] Su, J., Ng, D.T.K. and Chu, S.K.W. (2023) Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. *Computers and Education*: *Artificial Intelligence*, **4**, Article ID: 100124. https://doi.org/10.1016/j.caeai.2023.100124

[22] Kim, J. (2023) Leading Teachers' Perspective on Teacher-AI Collaboration in Education. *Education and Information Technologies*, **29**, 8693-8724. https://doi.org/10.1007/s10639-023-12109-5

[23] Ganguly, S. and Pandey, N. (2024) Deployment of AI Tools and Technologies on Academic Integrity and Research. *Bangladesh Journal of Bioethics*, **15**, 28-32. https://doi.org/10.62865/bjbio.v15i2.122

[24] Rane, N., Shirke, S., Choudhary, S.P. and Rane, J. (2024) Education Strategies for Promoting Academic Integrity in the Era of Artificial Intelligence and ChatGPT: Ethical Considerations, Challenges, Policies, and Future Directions. *Journal of ELT Studies*, **1**, 36-59. https://doi.org/10.48185/jes.v1i1.1314

[25] Pratomo, A.B., Mokodenseho, S. and Aziz, A.M. (2023) Data Encryption and Anonymization Techniques for Enhanced Information System Security and Privacy. *West Science Information System and Technology*, **1**, 1-9. https://doi.org/10.58812/wsist.v1i01.176

[26] Rizvi, S.S., Bolish, T.A. and Pfeffer, J.R. (2017) Security Evaluation of Cloud Service Providers Using Third Party Auditors. *Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing*, Cambridge, 22-23 March 2017, 1-6. https://doi.org/10.1145/3018896.3025154

[27] Roshanaei, M., Olivares, H. and Lopez, R.R. (2023) Harnessing AI to Foster Equity in Education: Opportunities, Challenges, and Emerging Strategies. *Journal of Intelligent Learning Systems and Applications*, **15**, 123-143. https://doi.org/10.4236/jilsa.2023.154009