

人工智能在教学评估中的应用：综述与分析

The Application of Artificial Intelligence in Educational Assessment: Review and Analysis

廖俐丽¹, 杨九民^{2*}

¹ 华中师范大学

* a1425144880@mails.ccnu.edu.cn

【摘要】 人工智能（AI）在教育领域已呈现出蓬勃发展的态势，然而，在众多的研究之中，关于人工智能在教学评估中的应用并未得到充分的体现。本研究检索 220 篇 AI 评估学生学习的文献，经过筛选，最终对 56 篇文章进行综合分析。研究目的在于探讨人工智能评估在当下教育应用中的具体途径，以及其所面临的机遇与挑战。本研究期望通过深入分析人工智能评估在教育领域的应用以及挑战，为 AI 评估的发展提供有益的启示和帮助。

【关键词】 人工智能；教育评估；AI 教育应用；大语言模型

Abstract: Artificial Intelligence (AI) has shown a burgeoning trend in the field of education. However, among numerous studies, the application of AI in teaching evaluation has not been adequately reflected. This study searched 220 documents on AI assessment of student learning, and after screening, conducted a comprehensive analysis of 56 articles. The purpose of the study is to explore the specific pathways of AI assessment in current educational applications, as well as the opportunities and challenges it faces. The study hopes to provide valuable insights and assistance for the development of AI assessment by deeply analyzing its application and challenges in the field of education.

Keywords: artificial intelligence, Educational evaluation, AI educational application, Large language model

1. 介绍

数字技术和计算机科学的进步 引领着社会走向一个技术时代,在这个社会中,机器的设计和发展逐步满足人类的需要,同时改变了教学的模式 (Espinosa, 2021)。本研究共检索 220 份相关论文,经过详细分析后,挑选了 56 份文献,这些文献展示了人工智能学习评估的应用与发展前景,同时也分析了 AI 评估所面临的挑战。

1.1. 人工智能评估

人工智能已经在各种领域中被研究了很长时间,包括在教育 (Memarian & Doleck, 2024)。然而,将人工智能等技术与教育直接联系起来的理论,特别是学习评估,还没有得到广泛的研究。一种流行的人工智能,被称为生成式人工智能提供 (Dwivedi et al, 2024)。它的发展与人类的自然语言处理系统和学习分析同步进行。

1.2. 痛点

ChatGPT 和其他人工智能工具的出现已经证明,机器学习系统 (AI) 可以通过培训或提示,以有效地完成评级任务及可反射式书写 (Jani, 2020),就像人类教师评分一样。近几年大量的人工智能研究涌现,但鲜少文章关注教育评估。人工智能为教育提供的可能性是巨大的,特别是对教育的辅导、评估和个性化学习。因此有必要促进人机协作辅助教学评估 (Fernández-Díaz, 2019)。

1.3. 研究问题

本研究把重点放在以下两个研究问题上:

RQ1:在回顾的研究中,人工智能评估已投入使用的类型有?

RQ2:人工智能评估在教育实践中的结果如何? 评估方法有何特征?

2. 方法

本研究综述遵循四阶段流程 (Liberati et al, 2009)、内容分析法对 56 篇文章进行分类和研究,以回答前文提出的两个研究问题。筛选流程包括以下步骤:

在中国知网、web of science、elsevier 数据库中都使用了关键词搜索,搜索“学习评估”、“人工智能”、“AI”、“AI 评估”、“智能学习辅导”等关键词。又通过“滚雪球”的方式检索到未使用到的搜索字符作为新的关键词进行第二轮检索。

下载所有文件的完整内容；识别重复的文献并剔除它们；阅览每篇文章的标题和摘要，决定是否相关；下载并阅读每篇文章的全文，决定是否相关；总结概述并分析结果；摘录并分类文献的实证研究类型；摘录并总结 AI 在教育领域的应用；探讨 AI 学习评估面临的挑战。

2.1. 检索过程

第一轮搜索获得 220 篇文章，剔除 26 篇重复文献后，剩余 194 篇。尽管多数研究涉及学习或形成性评估，但并非全部涵盖人工智能，72 篇因未明确使用 AI 技术被排除。另有 40 篇因颠倒 AI 与评估关系（即探讨评估促进 AI 发展而非 AI 用于评估）被弃用。此外，26 篇非教育领域研究也被剔除。最终，56 篇论文符合条件，用于系统综述，检索流程见图 1。



图 1 文献检索流程

3. 结果

3.1. 在回顾的研究中，人工智能评估已投入使用的类型有？

人工智能评估在多个教学领域展现出广泛应用。医学教学领域占比最高（39.3%），主要用于辅助医学生提升诊断准确率。语言类学习评估紧随其后（33.9%），随着技术的快速发展，机器学习不断完善，生成性人工智能能够有效识别语言和文本。使用生成性人工智能能够进行高质量的作文评估，并且能够为学生的书面评估提供形成性反馈。工程图类作业评估占 23.2%，研究多采用卷积神经网络（CNN）等深度学习模型，识别和分类 CAD 图纸元素，评估作业准确性，并通过机器学习算法提取关键特征，与标准图纸对比以检测误差。体育教学评估占比最低（3.6%），通过 AI 识别学生运动动态，评估其表现，适用于田径、游泳、篮球、足球等多个体育项目。各领域应用的可视化数据如图 2 所示。

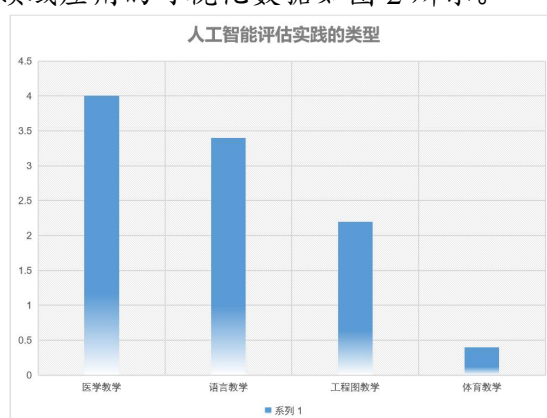


图 2 文献检索流程

3.2. 人工智能评估在教育实践中的结果如何？评估方法有何特征？

在语言学习评估方面，Pahi (2024) 研究显示，助教通过教学平台远程追踪学生学习进度，并结合人工智能进行人机协作，提供反馈，从而准确识别学生的学习进展或困惑，提升学习参与度和满意度。Pack (2024) 利用大语言模型自动评估英文论文，发现其与教师评估高度一致，有效减轻教师批改负担，推动评估模式变革。Awidi (2024) 则测试大语言模型是否能

按给定评分脚本评估论文，结果显示其不仅能提供语言反馈，还能通过自动评分节省教师时间，使其更专注于教学开发和个性化指导。学生通过即时反馈调整学习策略，提升成效。这些技术正革新传统教育模式，推动教育领域进步。

本研究在探讨中还发现绝大多数研究关注的是形成性评估，少部分是隐性评价的或者是有明确评价量表的。在本研究检索的文献中，有三份研究不使用人工智能进行英文学习形成性评估（Ulum, 2020）。其他以形成性评估为重点的论文认为：当教师拥有大量学生时，使用人工智能来完成评估任务可以为师生减轻教学负担（Goel, 2017），并为学生提供直接的学习反馈。多数研究通过 AI 自动标识学生作业实现评估，近半数文献详细描述了 AI 如何自动评分并提供反馈供教师使用。然而，少数研究如刘等学者的成果显示，AI 仅在任务结束时进行评分，未贯穿全过程（Liu, 2017）。

4.研究检索局限

本研究的工作存在一些局限性，我们的检索着重关注过程性学习评估和人工智能评估，但终结性性评估和机器学习等相关术语关注较少。此外，文章筛选时未充分考虑研究周期和被试量等因素，可能导致纳入研究周期短或被试量少的文献。未来，随着技术进步和数据积累，更多基于大规模样本的研究有望涌现，为领域提供更深入的洞见。我们将持续优化研究，吸纳更多样本丰富的文献，以增强综述的深度与广度。

5.人工智能评估的未来考虑

在传统教育与技术融合的教育环境中，学习评估的理论与智能评估实践之间往往存在一定的鸿沟。然而，随着人工智能等尖端技术的介入，对学生数据的评估过程得以显著简化，从而弥合了这一距离。

教育部科学技术与信息化新闻发布会强调，“发展数字教育，推进教育数字化与现代化是大势所趋、发展所需，也是改革所向。”在此背景下，教育智能化成为未来发展的重要方向。机器学习技术的引入不仅提升了学生作业评估的效率和准确性，还为教学方式改革提供了新的可能性。传统教学中，教师只能在课后批改作业，课堂上仅能点评个别学生的作业作为案例，而自动化评估系统使课堂更具灵活性。人工智能评估系统以其精准高效的特性，在教育实践中发挥了重要作用。将其融入智慧课堂，可显著提升教学互动与学习效率。教师在讲解知识点后，学生可立即完成随堂作业，并通过人工智能系统快速分析作业表现，实时获得反馈，及时向教师提问。这种人机协作的教学模式优化了反馈机制，缓解了教师资源有限的问题。学生可根据个性化反馈调整学习策略，实现更有针对性的学习，从而提升学习效果。

参考文献

- Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, 6, 100226.
- Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100002.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., & Ahuja, M. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Fernández-Díaz, E., Gutiérrez Esteban, P., & Fernández Olaskoaga, L. (2019). University-School Scenarios and Voices from Classrooms. Rethinking collaboration within the framework of an interuniversity project. *Journal of New Approaches in Educational Research*, 8(2), 79-95.
- Goel, A. K., & Joyner, D. A. (2017). Using AI to teach AI: Lessons from an online AI class. *Ai Magazine*, 38(2), 48-59.

- Jani, K. H., Jones, K. A., Jones, G. W., Amiel, J., Barron, B., & Elhadad, N. (2020). Machine learning to extract communication and history-taking skills in OSCE transcripts. *Medical Education*, 54(12), 1159-1170.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine*, 151(4), W-65.
- Liu, M., Wang, Y., Xu, W., & Liu, L. (2017). Automated scoring of Chinese engineering students' English essays. *International Journal of Distance Education Technologies (IJDET)*, 15(1), 52-68.
- Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, 2(1), 100040.
- Ocaña-Fernández, Y. *Inteligencia artificial y Sus Implicaciones En La Educación Superior. Propósitos Y Represent.*, 7 (2),(2019).
- Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence*, 6, 100234.
- Pahi, K., Hawlader, S., Hicks, E., Zaman, A., & Phan, V. (2024). Enhancing active learning through collaboration between human teachers and generative AI. *Computers and Education Open*, 6, 100183.