

人工智能在教育测评领域的应用与研究

李若仪¹, 宋亚光^{2*}

1. 伦敦大学学院, 英国 伦敦 WC1H 0AL

2. 山东第一医科大学, 山东 泰安 271016

DOI: 10.61369/SDME.2025240007

摘要: “教育测评”，实则包含两大重要组成部分，分别为“测量”与“评价”。随着人工智能时代的来临，人工智能正深刻改变着教育测评的形式并丰富了测评内容，为教育测评工作的创新发展提供了强有力的技术支撑。人工智能在教育测评领域的应用场景主要集中在机器命题、机器答题与自动评分三方面。本文一一对这些应用场景展开详细介绍并对人工智能在教育测评领域的应用提出合理化建议，以期尽快实现智能化测评目标，推动我国教育事业的蓬勃发展。

关键词: 人工智能；教育测评；应用；建议

Application and Research of Artificial Intelligence in the Field of Educational Assessment

Li Ruoyi¹, Song Yaguang^{2*}

1. University College London, London UK, WC1H 0AL

2. Shandong First Medical University, Tai'an, Shandong 271016

Abstract : "Educational assessment" actually consists of two important components, namely "measurement" and "evaluation". With the advent of the artificial intelligence era, artificial intelligence is profoundly changing the form of educational assessment, enriching its content, and providing strong technical support for the innovative development of educational assessment work. The application scenarios of artificial intelligence in the field of educational assessment mainly focus on three aspects: machine-generated test questions, machine answering, and automatic scoring. This paper introduces each of these application scenarios in detail and puts forward reasonable suggestions for the application of artificial intelligence in the field of educational assessment, aiming to realize the goal of intelligent assessment as soon as possible and promote the vigorous development of China's education industry.

Keywords : **artificial intelligence; educational assessment; application; suggestion**

引言

教育测评在教育改革与发展中始终处于关键地位，是检验教育质量的重要手段。《深化新时代教育评价改革总体方案》重点强调应将教育测评与信息技术紧密结合起来，以科技驱动教育测评改革与创新，目的是将教育测评的导向、诊断与激励作用发挥得淋漓尽致^[1]。然而，当前很少有研究聚焦人工智能+教育测评，尤其缺少对操作路径的研究与实践。而在笔者看来，如何以人工智能技术为抓手，构建教育测评新范式，为学生高效学习与全面发展提供实用性帮助，满足社会与国家发展需求，这应该是教育测评理论研究者与实践者密切关注的重点问题。

一、人工智能在教育测评领域的应用场景

(一) 机器命题

传统的命题主要由学科专家或者专业命题人员负责。他们首先会精准定位考试目标，立足目标，再设计试题，因此，可能带有极强的主观色彩。众所周知，教育测评的整体质量与命题质量

息息相关。通常来讲，一份合格的试卷应涵盖所有待考核知识点的代表性样本，同时，试卷难度得当，尤其应与测试目的相匹配。一般情况下，选拔性考试要比达标性考试的难度更高。随着在线学习系统的出现以及自适应考试逐渐走向教育领域，试题数量的需求大幅度增长。与传统纸笔考试不同的是，自适应考试需要按照学生能力水平匹配差异化题目，这可能导致题目需求量成

通讯作者：宋亚光，女，山东第一医科大学，山东第一医科大学心理健康教育中心主任，医学硕士、临床医师、心理治疗师、国家高级心理咨询师，从事心理特色实践育人、思想政治教育及心理健康教育、咨询、危机干预、课程建设等工作，Email: 18263806668@163.com。

倍增长^[2]。不仅如此，信息化时代的在线测试频率越来越高，这也对试题数量提出了增长需求。正因如此，传统命题方式的弊端也日渐明显，比如成本高、周期长且具有较高的出错率。反之，机器命题能更好地规避这些问题，同时还能大幅度提高命题效率，有效避免试题被泄露的风险。也正因如此，机器命题的发展迅速且应用范围广泛。

机器命题的模式大致可以分为两种，分别为强理论模型与弱理论模型。前者指的是基于扎实的认知理论命题。以数学命题为例，首先，测量学专家可以深入分析考生答题数据并估算各步骤的难度与权重；其次，由计算机自动替换题目中的部分元素。与母题相比，由测量学专家+计算机生成的试题变化多样且难度可控^[3]。而在现如今的教育领域，机器命题大多采用弱理论模型，具体命题过程如下所示：首先，命题专家确定母题并对题目展开多层次结构分析，可以按照背景、内容、问题、辅助信息、选项等构建多层次题目模型，之后，确定替换内容。在此基础上，计算机第一步要做的是细致分析可替换部分的文本与问题难度，继而全面搜索语料库和数据库，选取合适内容替换先前确定的替换部分，最终生成与母题结构相似、难度相近的新题。

值得强调的是，尽管相较于传统命题，机器命题的效率高、成本低，但是其应用与发展依旧面临着一系列现实困境。第一，机器命题依赖专家对母题的选择与分析；第二，机器命题通常需要依赖模板才能生成新题，尤其无法根据题目特点生成具有极强针对性的题目；第三，简答题类似的开放性题目的标准答案设计需要依靠专门模型支持，因而，机器命题在这类题目中的应用可能会受到一定限制；第四，语料库质量是影响机器命题质量的重要因素之一^[4]。如果缺乏成熟的语料库，那么机器命题的应用效果会大打折扣。

（二）自动评分

当前，简答题、作文题等类似的开放性试题往往依赖的是评分员评分，操作起来极其耗时耗力。用机器评分代替评分员评分，能在有效节约人力与时间成本的同时大幅度提升测评效率。自动评分大致分为三个基本步骤，分别为文本转换、文本分析和机器打分^[5]。第一步，需要将语音或者手写的内容转换为计算机可识别、分析的文本。此环节离不开自然语言处理系统的强大支持。第二步，文本分析，常用的分析方法包括“隐含语义分析”与“人工神经网络”两种。前者可以将考生的答案转换为数字矩阵，通过计算其与标准答案矩阵之间的差距来分析文本，多适用于简答题。后者则更适用于类似作文的篇幅较长的答案。人工神经网络主要通过提取文本特征来构建评分模型。常见的文本特征包括关键词频、句式复杂度、连词使用情况等。第三步，机器打分。机器打分通常采用分类法、回归模型两种。分类法更适用于分值较低的题目，比如0~5分。具体来说，计算机将待评答案与已有评分样本精细化匹配并归入最接近的分数类别，以此来完成打分。回归模型则更适用于分值较高的题目，比如作文、实验题、课外拓展题等，具体打分过程为将文本特征作为自变量，通过训练好的回归方程计算出连续型得分。截至目前，诸如“Project Essay Grade”和美国ETS开发的“E-Rater”等较为

成熟的自动评分系统已经被应用于教育测评领域^[6]。新一代的自动评分软件除了能完成自动评分外还能根据评分模型为学生提供针对性的改进建议。

（三）机器答題

机器答題的突出优势表现为能大幅度降低试测成本。在建设题库的过程中，新题均需经过试测，在获取各项性能指标后方可投入正式考试。而传统的试测方式需要耗费大量的时间和人工成本，而且还可能存在试题泄露的风险。而机器答題技术除了能大幅度降低试测成本外还能有效降低试题泄露的风险。但是由于机器答題技术的复杂程度较高，截至目前，其规模化应用还未实现。但是值得一提的是我国的科大讯飞以及日本、欧美的多个研究团队正致力于相关课题的研究，希望能获得突破性成就^[7]。

二、人工智能在教育测评领域应用的合理化建议

（一）夯实感知技术基础，保障伴随式采集

数据挖掘、深度学习等人工智能技术在教育测评领域的应用，需要依赖智能化测评模型，而该模型的精确性与可靠性需要依靠的是大规模、高质量的数据集。然而，如何获取并持续积累高质量的教育数据，这是亟待解决的现实问题。当前，采集教育数据并非一件易事，一方面，频繁的数据采集可能会加重教师、辅导员的工作负担；另一方面，大量的纸笔测试成果无法为智能测评提供有价值的数据支持，因为纸笔测试很难被转化为可分析的数字信号。除此之外，教育数据采集还可能面临数字安全、隐私保护等一系列问题。为了针对性解决数据采集难题，各地教育主管部门应积极构建“政产学研”协同机制，通过主动对接科技企业，真正将先进的智能化感知技术逐步向教育领域引入，比如语音交互识别、课堂行为分析等，争取将学生在线上、线下、课堂、家庭等多场景中的学习行为、认知状态、情感体验等数据有机整合起来，实现连续性测评的目的，同时，构建高质量、多模态的教育数据库，以便为形成性测评、诊断性测评等测评模型的构建提供精准且丰富的特征输入^[8]。类似伴随式、无感知数据采集与汇集模式不仅不会增加教师、辅导员的工作负担，不干扰正常的教学秩序，而且能大幅度提升教育数据获取的便利性，在有效解决教育数据获取困难这一难题的同时还能为智能教育测评体系构建提供可靠的数据支撑。

（二）完善场景建模体系，促进规模化应用

当前，智能化测评的精准度优势极其明显，尽管如此，其信度和公平性可能会面临理论与实证的双重挑战。正因如此，当前大部分测评模型局限于在线自主学习场景中应用，因为这类场景的容错率较高，而尚未被广泛应用于传统纸笔测评或其他高利害的场景中。究其根本原因，在于传统纸笔测评可采集的行为特征有限，并且数据分散，这难以满足智能化模型对数据规模以及数据质量等提出的基本要求，继而可能会缩小其应用范围。为了推进人工智能技术与传统测评的深度融合，可以将形成性、诊断性等低利害应用场景作为切入点。以单次纸笔测评为例，智能技术可以分析学习者的作答过程数据，比如答题轨迹、时间分配

等，通过此方式，将他们认知结构中存在的薄弱环节充分挖掘出来，这样，有利于增强测评的诊断精度，提升教学的有效性，还能不过度依赖分数^[9]。而针对多次评测场景，则需要持续搜集、整合学习者的过程性数据，同时，构建学习过程与遗忘过程模型，旨在持续跟踪、监测学习者学习历程并对相关趋势展开细致化分析，尤其重点关注波动趋势，以构建基于庞大数据的学习风险预警机制，为早发现、早干预提供科学依据。

（三）培育师生数据素养，落实精准化教学

当前，广大师生的数据素养亟待提升，这可能会影响精准化教学目标的实现。站在教师的角度，因为他们难以根据测评结果优化完善教学过程，继而可能陷入“有数据、难应用”的现实困境。而站在学生的角度，他们无法根据数据反馈及时调整学习策略，可能会制约个性化学习目标的实现。为了有效破解困境，教师应主动更新教学理念，尤其应确立“数据驱动决策”的理念，更重要的是不断提升个人信息化教学能力，提高数据素养。学生

则应积极配合教师做出的决策调整并锻炼从复杂数据中快速识别自身薄弱之处的能力^[10]。当然，不论是教师还是学生，均应养成基于数据总结、反思的好习惯，为精准化教学和个性化学习目标实现提供科学依据，继而推动教与学的协同发展。

三、结语

综上所述，人工智能作为当今时代最具影响力、挑战性和催化力的战略性技术，其对各行各业的赋能作用正在逐步显现出来。将人工智能应用于教育测评领域，是重塑教育生态、深化教育改革发展的关键举措。通过本文的分析可知，人工智能在命题、评分、答题场景中的应用已初见成效，但是，在应用环节中涌现出来的问题同样不容忽视。未来，希望有更多学者和专家加入相关课题的研究，真正推动人工智能与教育测评领域的深度融合。

参考文献

- [1] 徐湘荷.国际视野下人工智能教育测评：应用、风险及规避[J].聊城大学学报（社会科学版），2025(4):139–147.
- [2] 张鑫森，朱青，蔡玉军，等.人工智能赋能体育教育测评的应用场景、风险隐忧与纾解方略[J].体育学研究，2024, 38(3):38–49.
- [3] 杨丽萍，辛涛.人工智能推动教育测评范式变革的机遇与挑战[J].中国考试，2022(10):13–21.
- [4] 席小明.教育测评和学习领域的人工智能技术评估[J].中国考试，2021(5):56–62, 71.
- [5] 杨华利，耿晶，胡盛泽，等.人工智能时代的教育测评通用理论框架与实践进路[J].中国远程教育（综合版），2022(12):68–77.
- [6] 彭波，王伟清，张进良，等.人工智能视域下教育评价改革何以可能[J].当代教育论坛，2021(6):1–15.
- [7] 赵国宏，盖延龙，马芸，等.“面向教学”测评”的智能教育软件评估指标体系[J].中国教育信息化，2024, 30(7):64–72.
- [8] 骆方，田雪涛，屠焯然，等.教育评价新趋向：智能化测评研究综述[J].现代远程教育研究，2021, 33(5):42–52.
- [9] 胡盛泽，杨华利，黄涛，等.基于情境－认知－技术三重视角的智能测评诊断框架、实践和应用[J].远程教育杂志，2023, 41(4):45–55.
- [10] 刘欧.人工智能时代的可持续技能测评[J].中国考试，2025(3):1–6.