

AI 赋能中华古诗文助学平台的效果评估体系构建综述

谢红亮 周景怡 唐湘琳 付嘉晴 邱 怡*

广西职业师范学院 广西 南宁 530007

【摘要】随着人工智能技术深度赋能教育领域,各类中华古诗文助学平台应运而生,其在激发学习兴趣、传承文化遗产方面展现出巨大潜力。然而,当前学界与产业界对于如何科学、全面地评估此类平台的有效性仍缺乏共识。本文旨在系统梳理信息管理与教育技术领域的经典评估理论,构建一个融合量化与质性方法的综合评估框架。该框架涵盖平台技术质量、学习成效、用户体验与情感态度、文化素养与认同四个核心维度,以弥补传统评估中重技术轻人文、重结果轻过程的不足,旨在为相关平台的科学迭代与行业健康发展提供理论依据与方法论支持。

【关键词】人工智能教育;评估框架;信息系统成功模型;学习分析;数字人文

DOI:10.12417/2705-1358.26.01.079

1 引言

在全球数字化转型与中华文化复兴战略的双重驱动下,融合人工智能技术的中华古诗文助学平台正逐步成为创新教育模式、传承文化遗产的重要载体。这些平台通过自然语言处理、知识图谱、多模态生成等前沿技术,实现了对古典诗文的智能化解读、可视化呈现与个性化导学,极大地拓展了学习路径与体验。然而,平台的繁荣发展背后潜藏着一个关键性问题:我们应如何确证其教育价值与社会效益?当前,许多平台的评估往往停留在用户数量、功能使用频率等表面指标,或仅关注知识点的短期记忆效果,而忽视了其在激发深层学习动机、培养审美能力与增强文化认同感方面的潜在作用。这种评估的片面性可能导致开发方向上的短视与偏差,无法真正支撑其长远的教育使命。

因此,构建一套科学、系统、兼具技术理性与人文关怀的评估体系,已成为推动该领域从“工具创新”迈向“价值创造”的迫切需求。本文立足于计算机科学与信息管理的交叉视角,旨在通过系统综述与理论整合,提出一个多层次、多方法的综合评估框架,以期为各类中华古诗文AI助学平台的效能验证与优化升级提供普适性的理论模型和实践指南。

2 构建评估体系的理论基础

构建评估体系的首要任务是奠定坚实的理论基础。本研究主要植根于信息系统成功模型与教育技术评估理论,并兼顾数

字人文领域的评估特殊性。

首先,信息系统成功模型为评估平台的技术属性与基础用户体验提供了经典范式。其中,最具代表性的是 DeLone & McLean 模型,该模型历经演进,确立了系统质量、信息质量、服务质量、使用意愿、用户满意度和净收益六个核心维度^[1]。将其映射至古诗文助学平台,“系统质量”关乎平台的稳定性与交互流畅度;“信息质量”则直接指向 AI 生成内容的准确性、权威性与逻辑性;“服务质量”包括技术支持与教学辅导的及时性。该模型确保了我们从技术接受的角度审视平台是否“可用且好用”^[2]。

其次,教育技术评估理论则聚焦于学习的核心目标。Kirkpatrick 的四层次培训评估模型(反应、学习、行为、结果)为此提供了结构化思路^[3]。在平台情境下,“反应”层次对应于用户的学习体验与满意度;“学习”层次关注知识、技能与态度的获取,可通过标准化的前后测进行衡量;“行为”层次考察学习成果在实际场景中的应用迁移;“结果”层次则评估其带来的长期影响,如学术表现的提升。此外,学习分析技术使得通过分析用户与平台的交互数据(如答题序列、视频观看时长、资源下载记录)来客观、精细地刻画学习过程成为可能,弥补了传统测验只能捕捉学习“快照”的不足。

然而,上述经典理论在衡量古诗文学习的独特价值——即文化传承与精神滋养时,均显露出局限性。数字人文领域的评

作者1: 谢红亮(2004年-),男,汉族,广西桂林,本科,广西职业师范学院,研究方向:应用电子技术教育;

作者2: 周景怡(2005年-),女,壮族,广西来宾,本科,广西职业师范学院,研究方向:应用电子技术教育;

作者3: 唐湘琳(2004年-),女,汉族,广西桂林,本科,广西职业师范学院,研究方向:应用电子技术教育;

作者4: 付嘉晴(2004年-),女,汉族,广西桂林,本科,广西职业师范学院,研究方向:应用电子技术教育;

作者5(通讯作者): 邱怡(1987年-),女,汉族,广西藤县,研究生,广西职业师范学院,研究方向:人工智能、教育技术、物联网技术;

项目信息:本项目由自治区级大学生创新创业训练计划项目资助,项目名称:广西职业师范学院2025年大学生创新创业训练计划项目《诗图语梦——中华古诗文AI助学平台》,项目级别:自治区级,项目类别:一般项目,项目编号:S202514684066。

估强调对“文化素养”、“审美感知”与“身份认同”等软性指标的关注。这要求我们的评估框架必须超越工具理性，引入能够捕捉人文教育本质的测量方法，实现技术有效性与人文有效性的统一。

3 综合评估框架的构建

3.1 平台技术质量维度

此维度是评估的基石，确保平台作为教育工具的基本效能。其核心在于衡量平台核心技术模块的可靠性与智能化水平。量化评估可包括：系统响应时间、故障率、并发处理能力等性能指标；以及采用系统可用性量表等标准化问卷收集用户主观易用性反馈。在信息质量方面，可引入专家评审团对AI生成的诗文释义、赏析、仿写等内容进行准确性、丰富性与文化契合度的评级。

3.2 学习成效维度

此维度直接关乎平台的教育核心目标，需从多时间尺度进行考察。短期成效可通过精心设计的知识前后测（如对诗文记忆、典故理解、格律辨识的测验）进行量化对比。中期成效则可借助学习分析技术，追踪用户在平台上的学习路径、任务完成率、练习正确率的趋势变化，从而动态评估其知识建构过程。对于更高阶的批判性思维与创作能力，则需采用质性方法，如对用户撰写的诗评、仿写作品进行内容分析，评估其理解的深度与创造性^[4]。

3.3 用户体验与情感态度维度

此维度是维持用户长期投入与忠诚度的关键。它关注用户在使用过程中的主观感受与情感反应。量化方面，可采用技术接受度模型问卷，测量用户对平台的有用性、易用性感知及其使用态度。此外，专门的学习兴趣量表可用于探测平台在激发学习古诗文内在动机方面的效果。质性研究则能提供更丰富的洞察，例如通过绘制用户体验旅程图，识别其学习过程中的情感高峰与低谷点，或通过开放式问卷深入了解吸引或阻碍其持续使用的具体因素。

3.4 文化素养与认同维度

此维度是评估框架的特色与升华，旨在衡量平台在实现文化传承这一终极使命上的成效。由于其内涵的抽象性，该维度需创造性综合多种方法。量化上，可借鉴并修订社会学、教育学中成熟的文化认同感量表，对其进行情境化改造后施测。质性研究在此维度扮演不可替代的角色：其一，可通过深度访谈，探寻用户对古诗文意境、诗人情怀与哲学思想的理解是否发生了深化与迁移；其二，可采用投射法，如让用户为指定诗作选择配乐或进行绘画创作并阐述理由，从而间接揭示其审美体验与情感共鸣；其三，可鼓励用户撰写反思日志，并通过文

本分析其文化身份认同的细微变化。

4 评估方法论：量化与质性的融合

本框架的有效实施，依赖于量化与质性研究方法的协同运用，以达成相互补充、相互验证的效果。

4.1 量化研究

量化研究的优势在于其客观性、可比较性与可推广性。它通过结构化的量表、标准化的测验和系统自动记录的行为日志，将复杂现象转化为可统计分析的数值数据。例如，通过分析数千名用户的SUS得分，可以稳健地判断平台整体的可用性水平；通过对比实验组与对照组的前后测成绩，可以在一定程度上归因平台对知识掌握的影响。

4.2 质性研究

质性研究的优势则在于其深度、情境性与解释力。它致力于理解现象背后的“为什么”和“如何发生”。当量化数据显示用户在某功能页面停留时间较短时，后续的焦点小组讨论或用户访谈可以帮助我们揭示其深层原因——是内容枯燥、界面混乱还是操作繁琐？在评估文化素养等复杂构念时，一份反思日志中所流露出的对家国情怀的真切感悟，其说服力远胜于量表上一个孤立的分数^[5]。

因此，最理想的评估策略是采用混合方法研究设计。例如，在大型用户群体中广泛发放量化问卷以把握总体趋势，同时有目的地选取典型用户（如高活跃度用户、中途流失用户）进行深度访谈，用鲜活的质性资料为冰冷的统计数据赋予生命与语境，从而形成更为全面、深刻的研究结论。

5 挑战与未来研究方向

尽管本框架力求全面，但在实际应用过程中仍面临诸多挑战，这也为未来的研究指明了方向。首要挑战在于软性指标的标准化与操作化。如何将“文化认同感”、“审美能力”等抽象概念转化为可观测、可测量的指标，并确保其信度与效度，是一项艰巨的任务。这需要计算机科学、教育学、心理学和古典文学领域的学者开展深入的跨学科合作。其次，是因果归因的复杂性。用户文化素养的提升是一个长期、受多因素影响的过程。如何在一个非受控的真实学习环境中，将观察到的积极变化明确归因于平台的使用，而非学校教学、家庭环境等其他因素，是评估研究需要持续攻克的方法论难题。最后，数据伦理与隐私保护不容忽视。学习分析涉及大量用户行为与个人数据的采集，必须在研究设计之初就明确告知用户并获得知情同意，建立严格的数据匿名化、脱敏和安全存储机制，恪守学术伦理的底线。

展望未来，评估技术的发展可与AI技术本身同步演进。

多模态学习分析是一个充满前景的方向，例如，通过整合眼动追踪、面部表情分析、语音情感计算等技术，可以在自然情境下无侵入地捕捉用户在欣赏诗画、吟诵诗文时的投入度与情感变化，为“沉浸感”和“审美体验”提供客观的生理与行为证据。此外，推动行业内部就核心评估指标形成共识，建立基准数据集与评估标准，将是促进整个领域健康、规范发展的关键一步。

6 结论

本研究针对当前中华古诗文 AI 助学平台评估实践中存在的碎片化与表层化问题，通过系统整合信息系统与教育技术领

域的经典理论，构建了一个涵盖技术质量、学习成效、用户体验与文化认同的四维综合评估框架。该框架的核心贡献在于突破了传统评估范式的局限，将人文精神的滋养与文化身份的建构置于与知识传授同等重要的位置，并倡导量化与质性方法的深度融合。本文旨在为平台开发者、教育工作者及研究者提供一个结构化的思考工具与实践蓝图。通过实施如此全面而深入的评估，我们不仅能更准确地衡量平台的即时效果，更能引导其开发方向，最终推动 AI 技术真正成为赋能中华优秀传统文化创造性转化与创新性发展的强大引擎，在数字时代焕发古老诗文的永恒魅力。

参考文献：

- [1] 陈子璇,刘和海.基于 D&M 的数字教育资源用户满意度研究[J].中小学电教,2025,(Z2):32-36.
- [2] 马新杰.基于 SOR 的生成式 AI 用户持续使用意愿研究[D].杭州电子科技大学,2025.
- [3] 路漫漫.基于柯式四层评价模型的现代学徒制协同育人效果评估[J].中国多媒体与网络教学学报(中旬刊),2025,(06):145-148.
- [4] 卢洪艳.职业院校教育数字化转型成熟度模型构建与应用研究[D].江西师范大学,2025.
- [5] 叶兰,周文琦,张培宁,等.人工智能素养能力框架比较研究[J].图书情报工作,2025,69(21):142-153.