

生成式人工智能素养量表 (GLAT) 研究述评

林俊毅^{1,*}, 龚雪霏¹, 武珍¹

1. 福建理工大学, 人文学院, 福建 福州, 350118

摘要: 生成式人工智能的快速发展对素养评估提出新挑战。现有研究多依赖自陈式量表, 且缺乏跨文化验证。基于此, 本研究采用 CiteSpace 工具与文献计量方法, 对 2021-2025 年间中英文文献进行可视化与定性分析。结果表明: 研究理论基础从“信息素养”向“AI 素养”演进; 热点聚焦于“学评融合”“人机协同”等主题; 现有量表虽形成“认知-技能-情感-伦理”四维结构, 但存在跨文化等值性验证不足、评估缺乏标准化等问题。本研究提出“认知-技能-情感-伦理-评估-跨文化”六维度量表, 为跨文化场景下的量表设计提供理论支撑。

关键词: 生成式人工智能量表; ALS; GLAT; 跨文化评估; 文献计量法

A Review of Research on the Generative AI Literacy Scale (GLAT)

Chunyi Lin^{1,*}, Xuefei Gong¹, Zhen Wu¹

1. School of Humanities, Fujian University of Technology, Fuzhou, Fujian, 350118

Abstract: The rapid advancement of generative artificial intelligence poses new challenges for literacy assessment. Existing research predominantly relies on self-report scales and lacks cross-cultural validation. Consequently, this study employs CiteSpace and bibliometric methods to conduct visualization and qualitative analysis of Chinese and English literature published between 2021 and 2025. Findings indicate the theoretical foundation has evolved from "information literacy" towards "AI literacy"; research hotspots center on themes such as "integration of learning and assessment" and "human-machine collaboration"; existing scales (e.g., GLAT) have established a four-dimensional framework encompassing "cognitive-skills-affective-ethical" domains, yet suffer from insufficient cross-cultural equivalence validation and a lack of standardized assessment procedures. This study proposes a six-dimensional model integrating "cognitive-skills-affective-ethical-assessment-cross-cultural" dimensions, providing theoretical support for cross-cultural scale design.

Keywords: Generative AI literacy; ALS; GLAT; Cross-Cultural Assessment; Bibliometrics

生成式 AI 技术的兴起标志着智能时代进入“人机协同”创造的新阶段^[1]。与传统素养不同, 生成式 AI 素养强调对生成内容的批判性评估、伦理判断及共创能力。OECD 和 UNESCO 等国际组织将“人机协同”与“伦理责任”列为核心要素, 但如何科学评估该素养仍面临挑战^[2, 3]。现有量表多聚焦一般性人工智能知识或自陈式态度, 存在社会期望偏差、样本单一及跨文化等值性不足等问题。为此, Jin 等开发了生成式 AI 素养评估测试 (GLAT), 通过客观任务设计与心理测量验证, 试图突破自陈式测评的局限, 在国际上具有代表性与引领性意义^[4]。本研究旨在回答三个问题: (1) 生成式 AI 素养研究的主要维度与理论基础是什么? (2) 生成式 AI 素养量表的心理测量特性如何? (3) 现有生成式 AI 素养评估工具的局限及改善方向是什么?

1 研究设计

1.1 文献来源

为确保研究客观性,本研究从CNKI和WOS数据库检索2021-2025年与文章相关的中英文文献。中文检索式设定为SU=“ChatGPT素养”OR“生成式人工智能素养”OR“人工智能素养评价体系”,英文检索式为TS=“ChatGPT literacy”OR“Generative AI literacy”OR“Generative AI literacy scale”。文献类型限定为Article与Review,初步获得中文文献666篇、英文文献819篇。经多轮筛选,最终纳入有效中文文献272篇、英文文献127篇。纳入标准包括:1.主题涉及AI素养概念界定、框架构建或评估测量;2.以量表开发、验证或应用为主要内容;3.提供实证数据。

1.2 研究方法

本研究采用CiteSpace6.3.R1版本作为核心分析工具,采用文献计量、定性内容分析相结合的研究范式。通过高频关键词共现及聚类分析,旨在探究GLAT研究领域的知识结构、演进脉络和前沿趋势。聚类模块值($Q>0.3$)和平均轮廓值($S>0.5$)均符合校验标准,确保结果有效性。

2 研究结果与分析

2.1 文献计量与关键词聚类分析

通过CiteSpace图谱分析,采用LLR算法进行聚类标签提取,选取关键词为提取对象。中文文献关键词聚类图谱(Modularity $Q=0.6232$, $S=0.8892$)显示出以“数字素养”、“信息素养”为根基,向“人工智能”、“核心素养”蔓延的扩散性结构,表明国内研究仍处于传统素养与新型素养的碰撞与融合期,更注重将新概念纳入现有教育框架。

反观英文文献关键词聚类图谱(Modularity $Q=0.5643$, $S=0.8101$),则呈现出以“Trust”、“Explainable AI”、“Human-AI Collaboration”为核心的“技术-伦理-实践”三重维度紧密交织的形态,凸显出国际研究对技术可信度及协同机制的前沿探索,更聚焦于解决人机协同带来的全新挑战。

2.2 主题定性分析

为较为全面地反映GLAT研究领域的知识结构、演进脉络和前沿趋势,本文基于关键词共现和聚类可视化分析,将核心主题归纳为五大类(见表1)。

表1 基于CiteSpace聚类的研究主题定性分析

Table 1 Qualitative Analysis of Research Themes Based on CiteSpace Clustering

分类名称	核心关键词	内涵阐释	文献占比	总文献数
可信人工智能技术	可解释AI、生成式AI	可解释、可问责、可验证、可伦理、可泛化的智能系统技术范式	21%	83
素养框架的演进与构建	AI素养、信息素养、数字素养、核心素养	个体在智能时代需具备复合能力,是从传统信息技术向人机协作智能素养的转变与进化。	25%	99
教育应用与用户接受	持续使用意愿、接受度、知识工作、学习者态度	基于技术接受模型等领域,研究影响师生采纳AI工具的心理机制与行为结果,是技术实施的关键。	20%	80
人机协同	人机协同	关于重构教学流程、优化认知分工和增强人类智能的研究,是该领域的核心理念。	21%	83
教育评价体系革新	学评融合、评价能力、教育评价	应对AI环境下学习过程与目标的变化,在构建新型评价体系以衡量高阶思维与核心素养。	13%	51

当前研究归为五大主题,“素养框架的演进与构建”为核心(25%),但侧重于概念探讨,缺乏足够的实证支持;“可信人工智能技术”(21%)也大多停留在理论层面;“人机协同”(21%)对协同机制探讨不充分;“教育应用与用户接受”(20%)过于集中在学生群体,几乎未探讨教师

接受机制；“教育评价体系革新”（13%）多集中于单一维度的理论探讨，缺乏整合性的实证工具。此外，各主题间缺乏有效的整合，导致理论构建、技术研发与教育实践之间存在差距。目前存在的这些不足严重制约了生成式AI素养研究的深入发展。

3 GLAT 的理论基础与维度结构

3.1 理论基础

GLAT 的开发基于多学科理论，其核心包括生成式 AI 能力模型、可信 AI 伦理框架和心理测量理论。Jin 等人提出的生成式 AI 能力模型将素养划分为知识、应用、评估和伦理四个核心维度，构成了 GLAT 的骨架^[4]。此外，欧盟可信 AI 伦理框架以公平、透明、问责指导 GLAT 伦理维度。在心理测量方面，GLAT 使用项目反应理论（IRT）和经典测试理论（CTT）来进行题目设计和效度验证，确保测试准确可靠。这些理论基础与国内学者（蔡迎春等、钟柏昌等）强调的技术认知、批判思维和伦理意识维度高度一致，表明该四维结构具有跨文化的理论共识^[5, 6]。

3.2 维度结构

GLAT 的结构由知识、应用、评估与伦理四个维度组成，题型采用多项、情境判断与模拟任务。其框架包含 20 道多项选择题，覆盖生成式 AI 的基本原理、输出评估与伦理问题。其他研究也提出类似结构，如 Zhang 等纳入“创造性应用”以强调创新潜力^[7]；Chiu 等则突出“技术适应性”，关注个体在快速发展的 AI 环境中的学习能力^[8]。总体而言，该四维结构涵盖了生成式 AI 素养的认知、实践与伦理层面，比传统 AI 素养框架更强调其独特能力要求，为系统评估提供了基础框架。

3.3 国内外生成式 AI 素养主要评估框架对比

对比国内外 GLAT 量表，揭示评估形式、维度、目标群体、文化侧重和评估焦点差异，为分析 GLAT 优劣势及应用场景提供参考（见表 2）。

表 2 国内外 GLAT 量表主要评估框架

Table 2 Primary Evaluation Frameworks for GLAT Scales Domestically and Internationally

量表名称	类型	题项数量	目标群体	样本量	因子数量	因子描述/核心维度	信度	参考文献
GenAI Literacy Questionnaire	自陈式	32	中国 EFL 教师	603	5	GenAI 知识 GenAI 使用 GenAI 设计 GenAI 评估 GenAI 伦理	$\alpha=0.823$	[9]
ChatGPT Literacy Scale	自陈式	25	大学生	822	5	技术熟练度、批判性评估、沟通能力、创造性应用、伦理能力	$\alpha=0.86$	[10]
GenAI Literacy Assessment Test (GLAT)	混合式	20	大学生	355	4	认知与理解、应用与使用能力、评估和创造、伦理	$\alpha=0.80$	[4]
Generative AI Competence Scale (GenAIComp)	自陈式	63	大学生	1000	5	信息与数据素养、沟通协作能力、数字内容创作、安全与伦理、问题解决能力。	$\alpha=0.823$	[11]
国际中文教师生成式人工智能素养评估量表	混合式	22	国际中文教师	343	5	认知与态度、选择与运用设计与制作、实践与应用、评价与反思	$\alpha=0.980$	[12]
研究生人工智能素养量表	自陈式	12	研究生	1226	4	认识和理解、应用、伦理、评估和创建	$\alpha=0.86$	[13]
生成式 AI 辅助量表	自陈式	15	中国大学生	765	3	情感、行为、认知	$\alpha=0.888$	[14]

大学生应用生 成式人工智能 学习能力评价 指标体系	自陈式	37	中国 大学 生	141	6	规划、提问、判断、调整、应 用、反思	$\alpha=0.93$	[15]
------------------------------------	-----	----	---------------	-----	---	-----------------------	---------------	------

系统对比国内外主流评估工具发现,目前的量表多采用自陈式形式,题目数量在 12 至 63 项之间。因子数量大都涵盖“认知理解”、“技术应用”、“伦理责任”及“评估创造”等维度。其中, GLAT 作为混合式评估的代表,其 CFA 模型指标 ($CFI=0.93$, $RMSEA=0.06$) 及信度均表现优异,在客观测量认知与应用能力方面具有显著优势。然而, GLAT 存在两重局限:一是跨文化适应性不足,西方主导的开发范式难以适应非英语语境;二是情感态度维度覆盖有限,忽视主观心理因素。从文化视角来看,国际量表侧重技术熟练度与创造性应用^[11],而国内的则更强调评价与反思及实践与应用^[12],这凸显了文化差异导致的评估偏差。

4 GLAT 的信效度分析

研究显示,现有生成式 AI 素养量表在信度与效度上表现稳健。信度方面,所综述的八个量表整体内部一致性良好: GenAI 问卷^[9]、ChatGPT 素养量表^[10]、GenAIComp^[11]、国际中文教师量表^[12]及大学生学习能力指标体系^[15]的 Cronbach's α 均 >0.90 ,显示优异一致性; GLAT^[4] ($\alpha=0.89$) 和研究生 AI 素养量表^[13] ($\alpha>0.80$) 虽略低,但仍处于“良好”至“优异”的范围内,符合心理测量要求。GLAT 各维度信度 α 为 0.76-0.85,重测信度 $ICC=0.82$,共同证明了其稳定性;学习态度量表^[14]三维度 $\alpha>0.75$,也达到可接受水平。效度方面,量表通过结构效度提供多维证据:结构效度上, GLAT 的 CFA 模型 ($CFI=0.93$, $RMSEA=0.06$) 强有力支持其四因子结构^[4],其他量表通过因子分析或验证性因子分析验证了其理论维度结构,题项中因子载荷均 >0.5 或 0.6;且大多数量表在开发过程中都经过了专家评审确保内容效度。综上,现有量表为生成式 AI 素养研究与应用提供了可靠工具基础。

5. 讨论

5.1 构建“自评与他评”融合的评估维度

现有素养评估工具多依赖自陈式量表,易受社会期许效应影响,造成自我效能感与实际能力之间的偏差。为此,本研究提出增设独立的评估维度,以在理论层面整合自评与他评。在生成式 AI 量表中,评估不仅包括通过客观测试衡量技术操作水平,还应涵盖对 AI 生成内容的批判性评价和对人机交互过程的反思能力。正如相关研究指出,评价与反思是素养的高阶表现^[12]。因此,引入评估维度可将素养考察从技能掌握延伸至技能应用的价值判断,也呼应了 Walter 等人强调的批判性思维在 AI 教育中的重要性^[1]。

5.2 确立本土与国际互通的跨文化维度

现有量表多基于西方教育语境开发,缺乏跨文化等值性验证,这也限制了评估工具的普适性。GLAT 虽在西方样本中验证了信效度^[4],但当应用于非英语母语群体时,其题目设计的文化逻辑可能产生偏差。蔡迎春等指出社会责任在不同文化中有不同解释框架^[5],缺乏跨文化维度,直接用西方量表会让测量失效。引入跨文化维度,非简单语言翻译,要显性纳入文化适应性指标。该维度要求个体在不同文化语境下理解、应用及调适 AI,识别文化偏见。补全此维度,才能解决现有量表文化适应性不足问题,为开发多语言基准量表提供理论支撑。

6 结论

GLAT 量表因子结构稳定、信效度高,能客观测评技能与表现,广泛用于技术培训、招聘及高等教育等领域。但它过度聚焦特定技能,有“场景特异性”。现有实证研究多开发定制化工具,虽

提升单次评估准确性,但通用性差。从人才培养宏观视角而言,需开发跨学科、跨职业的通用评估工具。GLAT 可高效检验认知与技术能力,却难评估伦理与态度等主观素养。未来需融合 GLAT 与 ALS 两大类型量表,构建“能力-态度”双维模型,形成评估闭环。

本研究探讨现有生成式人工智能素养量表,发现当前评估体系存在通用性不足、文化适应性有限及关键维度覆盖不全等问题。因此,需在现有框架中增加评估能力和文化适应性两个关键维度。本研究提出的六维度模型修正了自陈式量表主观偏差,回应了文化差异挑战。未来量表开发应基于此模型,融合 GLAT 客观测试范式与 AILS 情感维度,构建“能力-态度-文化”综合测评体系,推动通用型人工智能素养量表标准化与科学化。

参考文献

- [1] Walter Y. Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education[J]. International Journal of Educational Technology in Higher Education, 2024, 21(1): 15.
- [2] UNESCO (2021). AI and Education: Guidance for Policy-makers.
- [3] OECD (2021). OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots.
- [4] Jin Y, Martinez-Maldonado R, Gašević D, et al. GLAT: The generative AI literacy assessment test[J]. Computers and Education: Artificial Intelligence, 2025: 100436.
- [5] 蔡迎春,张静蓓,虞晨琳,等.数智时代的人工智能素养:内涵、框架与实施路径[J].中国图书馆报,2024,50(04):7184.
- [6] 钟柏昌,刘晓凡,杨明欢.何谓人工智能素养:本质、构成与评价体系[J].华东师范大学学报(教育科学版),2024,42(01): 71-84.
- [7] Zhang H, Perry A, Lee I. Developing and validating the artificial intelligence literacy concept inventory: An instrument to assess artificial intelligence literacy among middle school students[J]. International Journal of Artificial Intelligence in Education, 2025, 35(1): 398-438.
- [8] Chiu T K F, Chen Y, Yau K W, et al. Developing and validating measures for AI literacy tests: From self-reported to objective measures[J]. Computers and Education: Artificial Intelligence, 2024, 7: 100282.
- [9] Wang Y, Derakhshan A, Ghiasvand F. EFL teachers' generative artificial intelligence (GenAI) literacy: A scale development and validation study[J]. System, 2025: 103791.
- [10] Lee S, Park G. Development and validation of ChatGPT literacy scale[J]. Current Psychology, 2024, 43(21): 18992-19004.
- [11] Lee S C, Baby T, Vongvit R, et al. Development and validation of Generative AI Competence Scale (GenAIComp) among university students[J]. Technology in Society, 2025: 103059.
- [12] 范慧琴,王璐璐,曹芳,等.生成式人工智能赋能国际中文教师数字素养提升研究:评估框架、现状及建议[J].天津师范大学学报(社会科学版),2024,(03): 72-83.
- [13] 李艳,朱雨萌,孙丹,等.典型科研场景下生成式人工智能使用的差异性分析——学科背景与人工智能素养的影响[J].现代远程教育研究,2025,37(02): 92-101+112.
- [14] 王永亮,吴韩伟,Derakhshan Ali,等.生成式AI辅助中国大学英语学习者态度量表的开发与检验[J].北京第二外国语学院学报,2024,46(06): 30-45.
- [15] 刘敏,马舒琪,王娟娟.大学生应用生成式人工智能学习:模型构建与评价指标体系研究[J].现代教育技术,2025,35(11): 14-22.

^{1,*} 作者简介:林俊毅(1970-),男,经济法学与工商管理双博士,教授,研究方向:数字经济、创新创业教育、策略管理等。E-mail: lukecylin@qq.com。