

GenAI 时代人机交互能力评估：批判性思维的影响

Assessment of Human-AI Interaction Capability in GenAI Era: the Influence of Critical

Thinking

苏红利¹, 闫欣玉¹, 沈蓉², 李菲茗^{1*}

¹ 浙江师范大学教育学院

² 浙江师范大学心理学院

* feimingli@zjnu.edu.cn

【摘要】在生成式人工智能（GenAI）时代，评估人类与 GenAI 有效互动的能力至关重要。然而，现有评估工具主要依赖于自我报告量表，本研究设计并验证针对人类 - GenAI 互动行为过程的评估指标，并分析批判性思维等因素的影响。研究招募 121 名大学生借助 GenAI 完成两项情境任务，记录他们与 GenAI 互动的过程，依据量规量化其在各指标的得分。分析表明测评指标具有良好的信度和稳定的三维结构。互动能力在性别、学科背景、GenAI 使用频率、批判性思维水平和问题链逻辑结构上均存在差异。高批判性思维可抵消使用频率低的劣势，凸显了批判性思维在增强人类-GenAI 互动能力中的重要性。

【关键词】智能交互（HAI）能力；信息素养；批判性思维

Abstract: In the era of generative AI (GenAI), it is crucial to assess the ability of individuals to effectively interact with GenAI. However, existing assessment tools rely on self-report scales. This study designs and validates evaluation indicators targeting the human-GenAI interaction behavior process and analyzes the impact of factors such as critical thinking. The study recruited 121 university students to complete two scenario-based tasks using GenAI, recording their interactions with GenAI. Their performance on various indicators was quantified using a rubric. The analysis indicates that the assessment indicators exhibit high reliability and a stable three-dimensional structure. The interactive capability varies based on gender, academic background, frequency of GenAI usage, level of critical thinking, and the logical structure of the problem chain. High critical thinking skills can compensate for the disadvantage of low usage frequency, highlighting the crucial role of critical thinking in enhancing human-GenAI interaction ability.

Keywords: Human-AI Interaction Capability; Information Literacy; Critical Thinking

1. 引言

生成式人工智能相关应用正在深刻重塑信息产生、传播和利用的范式。然而，人类使用者在不了解 GenAI 系统内部的结构、工作原理和信息处理过程的前提下，被动地接受模型输出数据会带来认知依赖性风险(阮一帆,2023)。有研究者担心 GenAI 应用会弱化用户的逻辑演绎、事实推理等思维过程(邢瑶,2024)。同时，GenAI 的回答会带有一定的偏见和主观性，真假信息的混合会阻碍用户处理信息。因此，人类与 GenAI 交互过程中表征问题、获取和评估信息并获取理想解答的能力，即人类与 GenAI 的交互能力（HAI）尤为重要。

有研究强调人类-GenAI 交互能力是 AI 素养的一个重要的行为维度，但对其测量停留在自我报告的量表上（Yuan et al, 2024），无法基于真实的交互行为的证据评估真正的交互能力。因此，我们需要设计聚焦人类-GenAI 交互行为过程的评价指标和量规。与以往的媒介相比，GenAI 的拟人性（像一个权威的专家）会降低人类对信息的质疑，因而交互过程中批判性思维的介入至关重要。所以本研究从传统的信息素养入手，分析人类与 GenAI 交互过程中涉及的信息加工流程，形成 Human-AI 交互能力的基本测量框架，并基于批判性思维贯穿于信

息加工中各阶段的行为表现形成评价指标。

2.理论框架

2.1.智能交互能力的定义与内涵

人机交互能力随着技术的演进有着不同的表现形式，在信息时代主要指人们使用计算机获取和处理信息的能力；在数字时代主要指人们利用数字设备与技术获取和处理信息的能力；当智能技术出现，对人类提出了理解和使用智慧化信息产品与服务的要求。人类与 GenAI 交互的过程包括有计划地向生成式智能工具提问，从反馈的答案或生成的文本中学习、反思，产生新的问题，并且能够对 GenAI 产生的信息进行评价和筛选，最后整合 GenAI 的回答去解决实际问题。GenAI 的应用不像一般的搜索引擎的使用，它需要学习者提出连贯、清晰和明确的指令，才能获取更有价值的信息。除此之外，由于对 GenAI 生成的内容仍然可能存在谬误，学习者更加需要具备分析和评估 GenAI 生成信息质量的能力。因此，本研究认为人类与 GenAI 交互的能力（HAI-capability）是指与生成式人工智能产品交互的过程中，能够利用提示工程相关技术，获取、分析、评价相关信息，并能对他们进行整合的一种能力。

智能交互能力体现了 AI 素养中的行为层面，它需要的获取、分析、评价、整合信息的能力继承了信息素养的基因，而评估信息的清晰度、相关性、逻辑性等能力依赖于批判性思维。

2.2.批判性思维的内涵与评估

联合国教科文组织发布的《学生人工智能能力框架》指出：AI 时代学生的批判性思维至关重要。人类与 GenAI 工具交互中面对信息的未知性和不准确性，更需要批判性思维的赋能。

美国哲学协会曾召集 46 名专家通过多轮德尔菲法讨论达成共识：批判性思维技能维度包括解释、分析、评估、推理、说明，以及自我调节（Facione, 1990）。可见，批判性思维技能不仅与 HAI 能力在分析、评估维度完全重合，其推理和说明维度也与信息整合相关。更重要的是，批判性思维的自我调节维度为 HAI 能力的测量提供了关键启示——人类需在与 GenAI 交互中审查过程与结果，不断优化提问策略并寻找最佳解决方案。

2.3.问题提出

融合信息素养与批判性思维视角，HAI 能力被进一步定义为：人类与 GenAI 交互中批判性获取、分析、评估和整合信息，并通过审查过程调节行为的能力。为此，本研究旨在设计基于人-GenAI 交互行为过程的评价指标，并验证批判性思维及其他因素对 HAI 能力的影响。因此，本研究主要解决两个问题，分别是：

问题一：基于 HAI 能力的界定如何设计有效评估的方式？它的信度和效度如何？

问题二：HAI 能力受什么因素影响，是否与批判性思维有关？

3.HAI 能力的指标和量规的迭代生成

首先，HAI 能力框架源于信息处理的四个基础阶段（获取、分析、评估、整合），然而，我们发现在人与 GenAI 的交互中信息获取与分析始终相互交织，因为检索信息的动作总是伴随着对先前信息的分析，故最终确立三个维度的框架：信息获取与分析、信息评估和信息整合。

指标体系的建立历经四轮迭代：（1）基于理论界定，研究小组通过经验头脑风暴形成初稿；（2）采集大学生使用 GenAI 的交互数据，判断指标在真实场景中的适用性、敏感度及覆盖度；（3）整合反馈优化指标，通过测评任务进行验证；（4）结合评分区分度与被试访谈，调整指标合理性。最终删除非核心能力指标（如创新性提问）及低区分度指标（如“

明确阐述观点""总结 GenAI 回答")，形成包含 10 项核心指标的评估体系（详见表 1），量规见附录 A2。

4.任务设计和研究方法

4.1 任务设计

激发被试与 GenAI 交互时展现预设的能力指标并凸显个体差异是任务设计的核心问题。参考王宽（2008）等人提出的任务设计特点，遵循劣构性（采用无单一答案的两难问题）、真实性（选取社会与科学领域现实议题）及有限经验性（平衡被试认知起点与探索空间）三大原则。最终选定科学与社会情境各一例（附录表 A1）。

4.2 研究方法

被试和实验流程 本研究招募了 121 名在校本硕士生，每位被试借助 GenAI 完成两个情景任务，提交交互过程以及问题解答文本，并通过在线填答批判性思维问卷，总限时 45 分钟。

实验材料 本研究实验材料包括测评任务（附录 A1）、量规（附录 A2）、问卷。为减少文化差异的影响，本研究选取本土化的批判性思维量表(侯玉波,2022)，该量表共 17 道题目包括开放性、分析技能、运用三个维度，其总体和分维度信度分别为.89,.77,.91,.77。

5.研究结果

5.1 信度

本研究招募了 6 位评分者，培训后评分者一致性信度达到 0.76。每个样本由两位评分者评估，并用两者的平均分作为被试在各指标上的得分。基于所有指标计算本测评工具的总维度和三个分维度的内部一致性信度均在.75 以上。

5.2 效度

结构效度通过探索性因子分析(EFA)和验证因子分析(CFA)共同检验。首先，根据 EFA 的碎石图，确定有三个因素。最后三因子结构如表 1 所示：除了 U1（提出的有效问题的数量），其他指标载荷均高于 0.4（多数>0.6），说明指标与维度契合。其次，CFA 表明三因素模型拟合良好： $\chi^2/df=1.34$ ，CFI=0.975，TLI=0.965，RMSEA=0.053，SRMR=0.068。进一步分析显示三维度间呈低相关性（ $r<0.35$ ），表明信息获取分析、评估、整合是 HAI 能力中相对独立的成分。以上结果为本研究提出的三因素模型提供了有力的证据。

表 1 智能交互能力的维度、指标及其标准化因子载荷

维度	评分指标	标准化因子载荷		
F1:信息获取与分析	U1 有效提问数量	<0.4		
	U2 问题思维层级	0.734		
	U3 多角度提问	0.776		
	U4 对 GenAI 回答的追问	0.733		
	U5 基于 GenAI 的回答不断调整提示	0.503		
	U6 问题链有结构和逻辑关系	0.778		
F2:信息评价	U7 合理质疑次数及质量		0.768	
	U8 对 GenAI 的回答进行评价的质量		0.611	
F3:信息整合	U9 论证是否进行了综合分析			0.697

	U10 论证的逻辑强度			0.790
	U11 论证中论据的数量			0.854

效标效度选择批判性思维作为效标进行分析。相关性分析发现批判性思维与整体 HAI能力水平以及信息获取和分析维度之间存在显著相关性，分别为 0.317 和 0.355。这表明批判性思维较高的大学生更有可能通过有效的提问策略进行批判性获取和分析信息。

5.3 学生 HAI能力影响因素

本研究探讨的 HAI能力影响因素除了批判性思维倾向，也包括性别、学科、使用频率和提问逻辑结构。首先比较了不同性别、学科和 AI 使用频率的被试在 HAI能力及其三个维度的差异（附录 A3）；其次，比较了并列模式、全递进模式、先并列后递进这三类提问逻辑结构下的被试在 HAI能力及其三个维度的差异。分任务分析显示（附录 A4），先并列后递进模式在整体能力及“信息获取分析”维度表现最优。

运用多元回归模型进一步分析所有因素的综合影响，通过逐步回归法（Stepwise Regression）选择的最终模型（表 2 所示）发现除性别外其他因素对 HAI能力均有显著的预测作用。理科生相对于文科生、GenAI 高频使用者相对于低频使用者，提问逻辑结构为全递进或者先并列后递进的被试相对为于全并列的被试，批判性思维水平高的被试更倾向于具有高水平的 HAI能力。

表 2 线性回归结果图

	未标准化系数		标准化系数		
	B	标准错误	β	t	p
（常量）	-6.352	4.083		-1.556	.123
专业（理科=1）	1.624	.811	.175	2.001	.048
使用频率（高频=1）	3.862	1.010	.322	3.823	<.001
全递进（是=1）	2.436	1.000	.221	2.435	.017
先并列后递进（是=1）	2.044	.914	.201	2.236	.027
批判性思维	.082	.038	.185	2.151	.034

在批判性思维与 HAI能力相关的基础上，进一步分析批判性思维能否弥补使用经验对 HAI能力的影响。以批判性思维倾向为 X 轴，HAI能力为 Y 轴做的散点图（附录图 B1）显示：低频组的批判性思维对 HAI的影响斜率更高，且当批判性思维水平较高（如得分≥120）时，低频使用者（≤1 次/周）与高频使用者（多次/周）的 HAI能力趋同，表明高批判性思维可弥补低使用频率的局限。

6.结论

本研究通过设计基于人-GenAI 交互行为过程的评估指标，揭示了批判性思维在提升人类与 GenAI 互动能力中的关键作用。实验结果表明，所构建的三维能力框架（信息获取与分析、评估、整合）具有较高的信效度，并能够有效区分个体差异。批判性思维不仅与 HAI能力显著正相关，还能弥补 GenAI 使用频率不足的劣势，使低频使用者通过高水平思维技能达到与高频使用者相近的交互效能。此外，“先并列后递进”的提问逻辑模式和学科背景（理科生表现更优）也显著影响交互能力。这些发现为 GenAI 时代的教育实践提供了重要启示：在培养技术操作技能的同时，需强化批判性思维训练，以优化人机交互的问题解决效能。

参考文献

- 邢瑶,王帆 & 李雯洁.(2024).AIGC 环境下研究生复杂认知的生成演变——基于 ChatGPT 的“证据推理”.现代教育技术,34(04),47-59.
- Yuan, C. W., Tsai, H. Y. S., & Chen, Y. T. (2024). Charting competence: A holistic scale for measuring proficiency in artificial intelligence literacy. *Journal of Educational Computing Research*, 62(7), 1675-1704.
- 王觅&钟志贤.(2008).论促进知识建构的学习环境设计.开放教育研究,(04),22-27.

附录

因篇幅缘故附录放置链接 (https://pan.baidu.com/s/13Qg5ZuCzFPVkusSaLG4Ni_Q?pwd=8218) 中。