·医学信息研究·

国内外人工智能素养测评工具综述

林 燕 孙海霞 蒲思樾 陈振丽 钱 庆

（中国医学科学院/北京协和医学院医学信息研究所/图书馆 北京 100020）

〔摘要〕 **目的/意义** 综述国内外人工智能素养测评工具研究现状，为医学领域工具优化及实践应用提供循证支持。**方法/过程** 系统检索中国知网、万方数据、维普网、PubMed、Embase 和 Web of Science 等数据库，筛选人工智能素养量表的原始研究文献，提取量表开发特征、维度框架及测量学属性数据，总结基本特征与维度演进，并基于健康测量工具选择的共识标准评估方法学及测量学属性质量，进而形成等级推荐。**结果/结论** 共纳入 27 项测评工具，内容分析表明，量表维度框架及内涵随时间和技术迭代演进，从技能导向转向综合素养导向；质量评估显示，部分量表存在属性验证不全面、内容效度证据不足等问题。未来应加强跨文化调适、长期跟踪研究及动态维度优化，以推动智能素养测评工具的全面发展，更好地服务于医学人工智能素养教育，实现精准评估。

〔关键词〕 人工智能素养；量表；健康测量工具选择的共识标准；系统综述

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2025.10.004

A Systematic Review of Domestic and International Artificial Intelligence Literacy Assessment Tools

LIN Yan, SUN Haixia, PU Siyue, CHEN Zhenli, QIAN Qing

Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

〔**Abstract**〕 **Purpose/Significance** To systematically review the research status of artificial intelligence (AI) literacy assessment tools domestically and internationally, providing evidence－based support for tool optimization and practical application. **Method/Process** The databases including CNKI, Wanfang, VIP, PubMed, Embase, and Web of Science are systematically searched to identify original development studies of AI literacy scales. Data on development features, dimensional framework, and measurement properties are extracted. Basic features and dimensional evolution are summarized, and methodological quality as well as measurement properties quality are evaluated based on the consensus－based standards for the selection of health measurement instruments (COSMIN) guidelines, followed by grading recommendations. **Result/Conclusion** A total of 27 assessment tools are included. Content analysis reveals that the dimensional frameworks and connotations of the scales evolves over time and technological advancements, shifting from skill－oriented to comprehensive literacy－oriented approaches. Quality assessment indicates that some scales lack comprehensive validation of properties or sufficient evidence for content validity. Future efforts should focus on cross－cultural adaptation, longitudinal tracking studies, and dynamic dimensional optimization to advance the comprehensive development of AI literacy assessment tools, better serve medical AI literacy education, and enable precise evaluation.

〔**Keywords**〕 artificial intelligence (AI) literacy; scale; consensus－based standards for the selection of health measurement instru-

〔修回日期〕 2025－08－21

〔作者简介〕 林燕，硕士研究生，发表论文 1 篇；通信作者：钱庆，研究员，博士生导师。

〔基金项目〕 中国医学科学院医学与健康科技创新工程项目（项目编号：2021－I2M－1－057）；国家重点研发计划（项目编号：2022YFC3601005）；国家社会科学基金项目（项目编号：21BTQ069）。

ments（COSMIN）；systematic review

1 引言

随着人工智能（artificial intelligence，AI）技术的快速发展及其在医学科研、医学教育和临床实践中的广泛应用，AI 素养已逐渐成为医学相关研究和医疗人才培养的重要议题。联合国教科文组织（United Nations Educational，Scientific，and Cultural Organization，UNESCO）[1] 明确指出，AI 素养已成为每个公民不可或缺的能力。从信息素养到数字素养，再到 AI 素养，素养的内涵始终与技术发展同步迭代[2]。信息素养概念诞生于图书馆领域，1974 年美国信息产业协会主席 Zurkowski P G[3] 将其定义为“人们在解决问题时利用信息的技术和能力”。1997 年 Gilster P[4] 提出数字素养的概念为“正确使用和评估数字资源、工具和服务并将其应用于终身学习过程的能力”。AI 素养由 Burgsteiner H 等[5] 在 2016 年提出并将其定义为“理解人工智能驱动技术背后的基本知识和概念的能力”。

辨析信息素养、数字素养、AI 素养的理论缘起与内涵，多数学者认为三者之间存在延续与拓展。如郭亚军等[6] 认为信息素养和数字素养是 AI 素养的前提，但也强调技术演进促使个体素养要求发生显著变化。AI 技术具有自主性、学习能力和不可解释性等特点，正重塑人类与物理、数字和机器世界的互动[7]。Poria S 等[8] 指出，AI 素养与其他素养的差异主要体现在信息技术的使用与应用层面。国际公认的素养能力框架清晰地展现了三者的区别。例如，在技能层面，2000 年美国《高等教育信息素养能力标准》[9] 聚焦于信息的检索、评价与利用。2018 年 UNESCO 发布《数字素养技能全球参考框架》[10]，在信息素养基础上扩展了利用信息通信技术和互联网进行创造、创新、创业，特别是网络交互与数字内容创造的能力。2024 年 UNESCO 发布《学生人工智能能力框架》[11]，强调个体应依据自身需求选择适宜的 AI 工具，优化人机协作，并运用 AI 创造和解决问题。进一步聚焦 3 类素养共有的“交互”概念可以发现，信息素养与数字素养中的“交互”主要指与他人分享信息[12-14]；而 AI 素养的交互，更侧重于“人智交互”，即协助 AI 提升其“认知”能力，模仿人类思维自动转换或生成内容以解决问题[6]。用户与 AI 的互动通常基于社交逻辑，而非传统机器交互逻辑[15]。因此，有必要对 AI 素养内涵与框架进行更精确的界定和研究。

目前应用最广泛的 AI 素养概念是 Long D 等[16] 提出的“AI 素养是一组能力，使个人能够批判性地评估 AI 技术，与 AI 进行有效沟通和协作”。基于此，国内外学者积极探讨 AI 素养的理论框架。例如，张银荣等[17] 提出涵盖 AI 知识、AI 能力和 AI 伦理的三维素养模型；郑勤华等[18] 基于加涅的学习结果分类，认为 AI 素养由智能知识、智能能力、智能思维、智能应用、智能态度 5 个维度构成；Kim S 等[19] 则提出由 AI 意识、AI 知识和 AI 能力构成的框架。与此同时，各机构和组织也相继推出重要框架。除 UNESCO 发布的《学生人工智能能力框架》与《教师人工智能能力框架》外，美国数字承诺组织于 2024 年制定的《人工智能素养：理解、评估和使用新兴技术框架》[20] 明确了素养内涵，并为 AI 赋能学习及教育策略提供指导。2025 年 5 月经济合作与发展组织、欧盟委员会联合起草面向中小学教育的 AI 素养框架，旨在界定学生核心能力，并提供课堂场景示例和跨学科整合指导[21]。这些多层次、多角度的框架研究，突显了提升全民 AI 素养的战略意义。为有效评估素养培养成效，国内外学者近年来已开发多种 AI 素养测评量表。然而，现有量表在基本特征、维度内涵及测量学属性等方面存在一定差异。因此，有必要进行系统总结与分析，并据此为不同人群推荐科学可靠的测评工具，进而推动针对医学生等特殊群体的专业化测评工具的开发与验证。

2 资料与方法

2.1 检索策略

文献检索分国内和国外数据库检索。国内文献检索选择中国知网、万方数据、维普网，中文检索式为：（“人工智能素养”OR“AI 素养”OR“人工智能能力”OR“AI 能力”）AND（“量表”OR“问卷”OR“调查表”OR“测评工具”OR“评估工具”OR“自评工具”），检索语种为中文。国外文献检索选择 PubMed、Web of Science、Embase，英文检索式为：（“artificial intelligence literacy”OR“AI literacy”OR“artificial intelligence competence”OR“AI competence”）AND（“scale”OR“test”OR“questionnaire”OR“survey”OR“tool”OR“index”），检索语种为英文。检索时扩展同义词并追踪参考文献与相似文献。检索时间为建库至 2025 年 7 月 1 日。

2.2 纳入和排除标准

纳入标准：内容为人工智能素养测评工具的开发，包括编制、修订、汉化、简化等；至少报告一项 AI 素养量表的测量学属性，如信度、效度；可获得文献全文；文献类型为期刊论文或学位论文。排除标准：量表在文中仅作为结局指标的评估工具；未在目标人群验证；重复文献。

2.3 数据提取与文献评价

由两名研究者独立完成文献筛选和信息提取，如遇分歧则与第三方协商解决。资料提取内容包括：工具/量表名称、发表时间、研究地区、适用人群、量表维度及条目数、评分方法、样本量、量表测量学属性等。两名研究者依据健康测量工具选择的共识标准（consensus - based standards for the selection of health measurement instruments，COSMIN）进行独立评价，具体步骤如下。一是基于 COSMIN 偏倚风险清单评估工具方法学质量，结果分为“非常好”“良好”“模糊”以及“不良”4 个等级[22]。二是采用 COSMIN 质量标准评估工具测量学属性，结果包括“充分”“不充分”“不确定”3 个等级[23]。三是使用改良版推荐分级的评估、制订与评价（grading of recommendations assessment，development and evaluation，GRADE）方法对证据质量进行分级，形成 A、B、C 3 类等级推荐：A 类工具推荐使用；C 类工具不推荐使用；B 类工具质量评价介于 A 类与 C 类之间[24]。

2.4 文献检索结果和量表纳入

通过数据库检索，共获得文献 1 969 篇，其中中文 918 篇，英文 1 051 篇。剔除重复文献 155 篇、不符合主题的文献 1 628 篇，保留文献 186 篇（中文 54 篇，英文 132 篇），最终纳入文献 28 篇（中文 8 篇，含量表 8 个；英文 20 篇，含量表 19 个），见图 1。

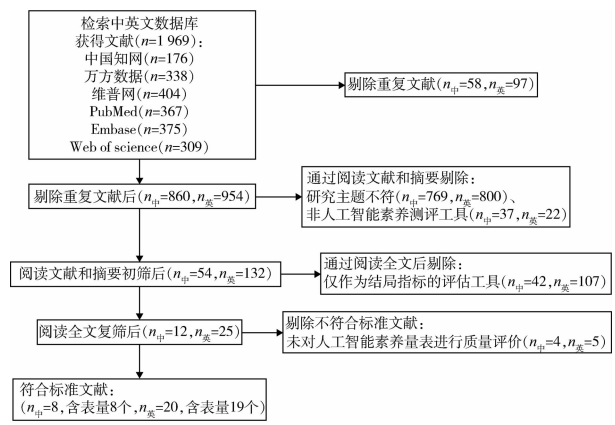

图 1 文献筛选流程

3 结果

3.1 人工智能素养量表基本特征

本研究共纳入 28 篇文献，涵盖 27 个 AI 素养量表，见表 1。总体而言，AI 素养量表开发发展迅速，研究时间集中于 2021—2025 年，学者群体主要来自中国、德国和韩国。量表适用对象以学生和教师为主，且重点关注学生群体的 AI 素养提升，涵盖小学、中学、大学不同阶段的智能素养能力测评。

量表维度数为 3 ~ 6 个，四维量表最为常见，各维度下设条目多为 7 ~ 10 条，量表总条目数介于 10 ~ 56 条之间。评分设置上，大多数量表采用易于理解和作答的李克特评分法，用以量化 AI 素养以

及主观态度，有部分学者为了减少主观因素干扰采取其他评分方式。例如，AILIT[25]、AI Literacy Tool[26]、AILIT－S[27]等采用二值评分法，每题明确对错答案，通过客观题型进行测评；AI Literacy[28]使用6级行为锚定量表，通过具体行为描述替代抽象表述，以达到减少主观偏差的目的。量表开发的样本量从几十到上千不等，其中多数研究的样本量超过200例，实证基础较充分，但也有少数研究的样本量不足100例。

表1 人工智能素养量表信息汇总

序号	发表时间	工具名称	国家/地区	应用人群	维度（个）/总条目（条）	评分方法	样本量（例）
1	2021.6	师范生智能素养自评工具[29]	中国	师范生	3/35	李克特量表（5级）	210
2	2021.12	MAIRS－MS[30]	土耳其	医学生	4/27	李克特量表（5级）	897
3	2022.5	AILS[31]	中国	一般人群	4/31	李克特量表（7级）	926
4	2023.4	高中生智能素养测评工具[32]	中国	中学生	4/35	李克特量表（5级）	362
5	2023.5	高中生智能素养问卷[33]	中国	中学生	3/27	李克特量表（5级）	208
6	2023.5	AILIT[25]	德国	大学生	31*	二值评分法	1 286
7	2023.6	大学生智能素养量表[34]	中国	大学生	4/55	李克特量表（5级）	519
8	2023.8	MAILS[35]	德国	一般人群	4/34	李克特量表（11级）	282
9	2023.8	AI Digital Literacy[36]	韩国	大学生	4/19	李克特量表（5级）	318
10	2023.9	SNAIL[37－38]	德国	大学生	3/31	李克特量表（7级）	415
11	2023.12	Chan & Zhou's Scale[39]	中国香港	大学生	4/22	李克特量表（5级）	405
12	2024.3	AISES[40]	中国台湾	一般人群	4/22	李克特量表（7级）	314
13	2024.5	公务员智能素养问卷[41]	中国	公务员	4/27	李克特量表（5级）	459
14	2024.5	中小学教师人机协同素养问卷[42]	中国	教师	3/37	李克特量表（7级）	1 017
15	2024.5	AILQ[43]	中国香港	中学生	4/32	李克特量表（5级）	363
16	2024.6	高校学生智能素养问卷[44]	中国	大学生	5/30	李克特量表（5级）	50
17	2024.6	ChatGPT Literacy Scale[45]	韩国	大学生	5/25	李克特量表（5级）	822
18	2024.9	AILS－CCS[46]	中国	大学生	4/14	李克特量表（5级）	546
19	2024.10	AI Literacy Scale[47]	中国	一般人群	5/15	李克特量表（5级）	302
20	2024.11	MAILS－CCS[48]	德国	大学生	10*	李克特量表（5级）	653
21	2024.11	AI Literacy[28]	拉脱维亚	教师	6/10	6级行为锚定量表	42
22	2024.12	Zhong & Liu's Scale[49]	中国	中学生	3/56	李克特量表（5级）	1 392
23	2025.1	AI Literacy Tool[26]	韩国	小学生	3/24	二值评分法	580
24	2025.3	AILST[50]	中国	教师	4/36	李克特量表（5级）	604
25	2025.3	AILIT－S[27]	德国、英国、美国	大学生	5/10	二值评分法	1 465
26	2025.4	TAICS[51]	中国香港	教师	6/24	李克特量表（5级）	434
27	2025.5	小学生智能素养量表[52]	中国	小学生	5/20	李克特量表（5级）	291

注：* 表示量表下未设置维度。

从研究成果发表的时间来看，2021年2个，2022年1个，2023年8个，2024年11个，2025年5个。2021年6月王欢[29]率先开发师范生智能素养自评工具，这是国内较早针对特定人群开发的AI素养量表，为师范生自我评估AI素养提供了量化工具。Karaca O等[30]于同年开发MAIRS－MS，这是首个用于评估医学生AI素养的量表，为医学教育领域及其他专业领域的AI素养测评提供了参考和借鉴。

智能化背景下，AI素养能力培养日益重要，相关测评工具的研发受到关注，量表数量显著增加。2023—2024年间，中国、德国等国家/地区的研究团队在此领域共开发19个量表。例如，Hornberger M等[25]的AILIT、Carolus A等[35]的MAILS以及Hwang H S等[36]的AI Digital Literacy等。由于文献纳入与排除标准的限制，本研究未涵盖此阶段发布在会议论文及预印本中的量表，例如，Weber P等[53]发表在会议论文集中的量表开发研究。

此后，AI素养量表开发持续受到学界高度关注，研究对象进一步细化。例如，中国学者与韩国

学者相继开发了面向小学生的专用量表[26,52]。同时，针对已开发量表的验证、更新与简化研究日益增多。例如 Laupichler M C 等[38]在开发 SNAIL 量表后，对其进行了进一步测试与改良；Koch M J 等[48]基于 Carolus A 等[35]开发的 MAILS 量表，推出了简化版本 MAILS－CCS；Hornberger M 等[25,27]则在其成熟的 AILIT 量表基础上，开发了简化版 AILIT－S，并在德国、美国和英国进行验证，表明简版量表同样具有良好信效度。

3.2 人工智能素养量表维度与内涵解析

开展 AI 素养测量前，须应用相应框架、模型以及理论对概念加以界定，以达到将抽象事物转化为具体构念的目的[54]。为了研究 AI 素养框架构成要素以及不同工具所测量的“人工智能素养”概念内涵是否一致，对纳入的 27 个量表的维度指标及内涵进行提取与整合，并进一步分析维度划分和测量内容。具体步骤如下：先对表述相似且内涵高度相关的维度进行归类合并，如将“智能知识”“智能认知”等统一归并为“AI 知识”，最终形成 AI 知识、AI 技能、AI 意识、AI 伦理、AI 思维 5 个核心测评维度；再针对同类维度的内涵差异，进一步梳理测量重点，以明确不同工具的测评方向和内容。每种工具涉及的维度和测评内涵，见表 2。

表 2 人工智能素养量表维度内涵分布

序号	AI 知识				AI 技能				AI 意识				AI 伦理			AI 思维			
	AI 概念通识	AI 技术原理	AI 设备知识	算法编程知识	AI 操作能力	AI 应用能力	AI 评估能力	AI 优化与创新	AI 态度和感知	AI 优劣势和预见	AI 兴趣和学习意识	信心和自我效能	AI 道德和法规	AI 社会影响	安全与隐私保护	算法思维	数据思维	批判思维	人机协同
1	√	√			√	√		√	√		√		√						
2	√	√	√		√	√			√	√			√						
3	√				√	√	√						√		√				
4	√				√	√		√	√	√	√		√	√	√	√	√		
5	√				√	√		√			√		√						
6	√	√				√			√	√	√		√	√					
7	√	√		√	√	√		√	√		√		√			√	√	√	
8	√	√			√	√		√	√	√	√			√					
9					√	√		√	√	√			√						
10	√	√			√	√				√			√	√	√			√	
11	√				√	√			√	√	√								
12									√		√	√							√
13		√	√		√	√			√				√		√	√			√
14	√				√								√				√	√	√
15	√				√	√			√		√	√	√		√				
16				√		√		√	√				√	√		√	√	√	√
17					√	√	√	√					√					√	√
18	√	√			√	√	√						√	√	√				
19					√	√			√	√		√	√					√	
20	√					√			√				√						
21	√	√					√						√	√	√	√			
22	√	√						√	√	√						√			
23	√				√	√				√	√		√						
24	√				√	√	√						√						
25	√	√				√			√				√						
26	√						√						√						
27	√	√				√	√	√	√		√		√		√	√			

注：表中序号与前文表 1 相同，按量表发表时间顺序编号。

通过对比发现，现有量表维度框架既存在共性也呈现差异。共性在于 AI 素养框架具有同一性，这或许是因为理论同源，借鉴并沿用了相同的理论框架。如 Ma S 等[46]开发的 AILS－CCS 模型框架借鉴了 Wang B 等[31]开发的 AILS，两者维度相同。差异则体现在不同工具对 AI 素养的测量重点不同。早期量表多聚焦基础能力测评，主要涵盖知识和技能的掌握和熟练程度，如师范生智能素养自评工具[29]包括智能知识、智能能力、智能意识 3 个维度，侧重对 AI 知识理解、AI 技能掌握、AI 可能引发的伦理问题等的评估。但随着人智交互的发展，学者开始探讨人工智能素养的其他社会情感和高阶认知维度，AILQ[43]的情感、行为、认知和伦理 4 个维度更注重个人对 AI 的感知或协作，着眼于评

估社会情感素养。

进一步分析发现，AI 素养的测量内涵呈现动态演进趋势，其维度定义随技术迭代与研究深化不断拓展。AI 知识维度早期多为数据统计、分析、技术原理、机器学习等基础知识熟悉程度的测量，逐步扩展至生成式 AI（如 ChatGPT）的应用原理及产品特性认知。AI 意识维度从最初的对 AI 的态度、兴趣和学习意愿，进一步延伸至自我效能考量。如 Wang Y Y 等[40]专门开发 AISES 用于测量个人使用 AI 技术/产品的信心和自我效能感。基于越来越多学者对 AI 的辩证思考，AI 思维维度在基础的计算思维、编程思维上增加批判性思维和人机协同思维考量，如中小学教师人机协同素养问卷[42]通过高阶数智思维条目，测评教师在人机协作中的决策与创新能力。其中数智思维指在理解数智融合的基础上，在人机协同中主动发现问题、批判评估 AI 结果并作出合理决策的思维能力。上述演变表明，量表维度内涵从基础到高阶递进，AI 素养测评已从基础技能评估转向对高阶能力的系统性考察，反映出技术发展对素养内涵的持续重塑。

3.3 基于 COSMIN 的人工智能素养量表质量评价

28 项研究均未报告工具的可靠性、测量误差及效标效度相关信息，其他方法学及测量学属性质量评价，见表 3。

表 3 纳入量表的方法学及测量学质量评价

量表名称	PROM	内容效度	结构效度		内部一致性		假设检验	反应度
	结果	结果	指标	结果	Cronbach's α 系数	结果	结果	结果
师范生智能素养自评工具[29]	D	D	EFA、CFA、TLI = 0.847、CFI = 0.856、RMSEA = 0.07、SRMR = 0.061	V/+	0.953	V/+	NR	NR
MAIRS-MS[30]	D	D	EFA、CFA、CFI = 0.938、RMSEA = 0.094、SRMR = 0.057	V/+	0.87	V/−	NR	NR
AILS[31]	D	D	EHA、CFA、TLI = 0.99、CFI = 0.99、RMSEA = 0.01、SRMR = 0.03	V/+	0.83	V/+	V/+	NR
高中生智能素养测评工具[32]	D	D	NR	NR	0.983	A/?	NR	NR
高中生智能素养问卷[33]	D	D	EFA	A/?	0.929	V/+	NR	NR
AILIT[25]	D	D	CFA、TLI = 0.964、CFI = 0.966、RMSEA = 0.026、SRMR = 0.039	V/+	0.82	A/?	V/+	NR
大学生智能素养量表[34]	D	D	EFA、CFA、TLI = 0.996、CFI = 0.997、RMSEA = 0.028	V/+	0.965	V/+	V/+	NR
MAILS[35]	D	NR	CFA、CFI = 0.926、RMSEA = 0.057、SRMR = 0.079	V/+	*	A/−	V/+	NR
AI Digital Literacy[36]	D	NR	EFA、CFA、TLI = 0.90、CFI = 0.91、RMSEA = 0.07	V/−	*	A/+	NR	NR
SNAIL[37]	D	D	EFA	A/?	*	A/+	V/+	NR
SNAIL 德语版[38]	D	NR	NR	NR	NR	NR	NR	V/+
Chan & Zhou's Scale[39]	D	NR	CFA	V/?	*	A/+	NR	NR
AISES[40]	D	D	EFA、CFA、TLI = 0.93、CFI = 0.941、RMSE = 0.079、SRMR = 0.071	V/+	0.958	V/+	NR	NR
公务员智能素养问卷[41]	A	D	EFA、CFA、TLI = 0.918、CFI = 0.829、RMSEA = 0.041	V/+	0.959	V/+	NR	NR
中小学教师人机协同素养问卷[42]	D	NR	EFA、CFA	V/?	0.981	V/+	NR	NR
AILQ[43]	D	D	EFA、CFA、TLI = 0.91、CFI = 0.92、RMSEA = 0.06、SRMR = 0.06	V/+	0.93	A/?	V/+	NR
高校学生智能素养问卷[44]	D	D	NR	NR	NR	NR	NR	NR

续表 3

量表名称	PROM	内容效度	结构效度		内部一致性		假设检验	反应度
	结果	结果	指标	结果	Cronbach's α 系数	结果	结果	结果
ChatGPT Literacy Scale[45]	D	D	EFA、CFA、TLI = 0.92、CFI = 0.93、RMSEA = 0.06、SRMR = 0.05	V/+	*	A/+	V/+	NR
AILS-CCS[46]	D	D	EFA、CFA、CFI = 0.941、RMSEA = 0.076	V/-	*	A/+	V/+	NR
AI Literacy Scale[47]	D	D	EFA、CFA、CFI = 0.976、RMSEA = 0.043	V/+	0.755	V/-	V/+	NR
MAILS-CCS[48]	D	NR	CFA、TLI = 0.936、CFI = 0.958、RMSEA = 0.090、SRMR = 0.038	V/+	NR	NR	NR	NR
AI Literacy[28]	D	D	NR	NR	0.872	A/?	NR	NR
Zhong & Liu's Scale[49]	D	D	EFA、CFA、CFI = 0.863、RMSEA = 0.055、SRMR = 0.060	V/+	>0.9	A/?	V/+	NR
AI Literacy Tool[26]	D	A	NR	NR	0.826	V/+	NR	NR
AILST[50]	D	D	EFA、CFA、TLI = 0.945、CFI = 0.982、RMSEA = 0.063	V/+	0.889	V/+	V/+	NR
AILIT-S[27]	D	NR	CFA、TLI = 0.977、CFI = 0.982、RMSEA = 0.021、SRMR = 0.040	V/+	0.061	A/-	V/+	NR
TAICS[51]	D	D	CFA、CFI = 0.96、RMSEA = 0.054、SRMR = 0.031	V/+	*	A/+	V/+	NR
小学生智能素养量表[52]	D	D	EFA	A/?	0.886	A/?	NR	NR

注：PROM 表示患者报告结局测量工具。结构效度指标中，EFA = 探索性因子分析，CFA = 验证性因子分析，TLI = 塔克-刘易斯指数，CFI = 比较拟合指数，RMSEA = 近似误差均方根，SRMR = 标准化残差均方根。方法学质量中，V = 非常好，A = 良好，D = 模糊，I = 不良；测量学属性中，+ = 充分，- = 不充分，? = 不确定。* 表示虽然报告了各维度的内部一致性，但未报告量表整体的 Cronbach's α 系数。NR 表示未报告该项内容。

3.3.1 纳入研究的方法学质量和测量学属性评价 （1）量表开发质量。28 项研究均在量表开发方面清晰描述了所测构念，大多基于文献综述形成理论框架，只有公务员智能素养问卷[41]和高校学生智能素养问卷[44]基于扎根理论构建模型。其中，公务员智能素养问卷[41]开发过程中有明确的访谈提纲和记录，并双人录入信息，满足“良好”的评价要求。其余 27 项研究缺乏对量表开发过程的详细描述，故评为“模糊”。（2）内容效度。在报告内容效度的 21 项研究中，AI Literacy Tool[26]通过两轮德尔菲专家验证并采用内容效度比（content validity ratio，CVR）量化专家共识，评为“良好”。其余 20 项研究的内容效度评价均为“模糊”。具体原因如下：5 项研究[29,32-34,44]仅通过专家函询进行评价，缺乏量化指标或对条目相关性、全面性及理解性等维度的具体说明；2 项研究[25,28]的专家咨询人数少于 4 人；小学生智能素养量表[52]虽对测评对象进行了预调查，但未提及是否经过专家咨询；其余报告内容效度的研究虽通过专家咨询法征集意见，但缺少对上述具体维度的详细描述。（3）结构效度。共 23 项研究评价量表的结构效度。其中 14 项[29-31,34,36,40-43,45-47,49-50]进行了探索性因子和验证性因子分析，且样本量超过条目数 5 倍，方法学质量评为“很好”。另有 6 项研究[25,27,35,39,48,51]采用验证性因子分析，评为“很好”。而其余 3 项[33,37,52]只进行探索性因子分析，评为“良好”。就量表的测量学属性而言，23 项研究中的 2 项[36,46]因拟合优度不满足 CFI 或 TLI 大于 0.95、RMSEA <0.06 或 SRMR <0.08，评为“不充分”，5 项[33,37,39,42,52]由于未报告量表的拟合优度情况，被评为“模糊”，其余 16 项[25,27,29-31,34-35,40-41,43,45,47-51]拟合优度良好，测量属性“充分”。（4）内部一致性。共 25 项研究验证了量表的内部一致性。其中 11 项[26,29-31,33-34,40-42,47,50]计算了量表整体及各维度的内部一致性，方法学质量评为“很好”。另有 7 项[25,27-28,32,43,49,52]仅计算了量表整体内部一致性，方法学质量评为“良好”。剩余 7 项研究[35-37,39,45-46,51]

计算了量表各维度的内部一致性，但并未计算量表整体内部一致性，评为“良好”。对于量表的测量学属性而言，25 项研究中的 14 项[29,31,33-34,36-37,39-42,45-46,50-51]验证了量表各维度的 Cronbach's $\alpha \geq 0.70$，测量学属性评为“充分”。较为特殊的是，AI Literacy Tool[26]采用 KR-20 系数作为信度指标，其值为 0.826，因此同样评为“充分”。而 4 项研究[27,30,35,47]存在 Cronbach's $\alpha < 0.70$ 的情况，评为“不充分”。此外，还有 6 项[25,28,32,43,49,52]未报告具体数值，评为“模糊”。(5) 假设检验。在报告假设检验的 13 项研究中，两项研究[34,47]采用结构方程模型，其余 11 项[25,27,31,35,37,43,45-46,49-51]采用验证性因子分析。由于对各亚组特征描述充分，且分析结果与理论假设相符，方法学质量评为“很好”，测量学属性评为“充分”。(6) 反应度。仅 SNAIL 德语版[38]通过评估智能教育课程参与前后的人工智能素养变化验证了反应度。其方法学质量因研究设计严谨被评为“很好”，且测量学属性因结果与假设一致被判定为“充分”。其余量表未报告反应度分析。

3.3.2 纳入工具的证据等级评价及推荐意见 本研究采用改良版 GRADE 证据分级方法，具体步骤如下：假设证据均为“高质量”，依次从偏倚风险、不一致性、不精确性、间接性 4 个方面进行降级，形成“高、中、低、极低”4 个等级，再根据结果形成 A、B、C 3 类等级推荐[55]。在证据分级过程中，对于偏倚风险，如果方法学质量“良好”无须降级，“模糊”降一级，“不良”则降两级；如果量表的多项研究存在不一致结果降一级；评估不准确性，如果量表开发过程中的样本量在 50~100 之间降 1 级，小于 50 降 2 级；如果证据视为间接证据则进一步降级。

综合评价结果表明，27 个量表均未在内容效度、结构效度及内部一致性 3 方面同时达到“充分”标准，因此无 A 级推荐。其中，高校学生智能素养问卷[44]与 AI Literacy[28]因样本量不足（≤50）、内容效度低及属性报告缺失被评定为 C 级，其余量表均为 B 级，见表 4。

表 4 纳入量表的证据等级评价及推荐

量表名称	内容效度	结构效度	内部一致性	假设检验	反应度	评价结果
师范生智能素养自评工具[29]	中	高	高	NR	NR	B
MAIRS-MS[30]	中	高	高	NR	NR	B
AILS[31]	中	高	高	高	NR	B
高中生智能素养测评工具[32]	中	NR	高	NR	NR	B
高中生智能素养问卷[33]	中	高	高	NR	NR	B
AILIT[25]	中	高	高	高	NR	B
大学生智能素养量表[34]	中	高	高	高	NR	B
MAILS[35]	NR	高	高	高	NR	B
AI Digital Literacy[36]	NR	高	高	NR	NR	B
SNAIL[37-38]	中	高	高	高	低	B
Chan & Zhou's Scale[39]	NR	高	高	NR	NR	B
AISES[40]	中	高	高	NR	NR	B
公务员智能素养问卷[41]	中	高	高	NR	NR	B
中小学教师人机协同素养问卷[42]	NR	高	高	NR	NR	B
AILQ[43]	中	高	高	高	NR	B
高校学生智能素养问卷[44]	低	NR	NR	NR	NR	C
ChatGPT Literacy Scale[45]	中	高	高	高	NR	B
AILS-CCS[46]	中	高	高	高	NR	B
AI Literacy Scale[47]	中	高	高	高	NR	B
MAILS-CCS[48]	NR	高	NR	NR	NR	B
AI Literacy[28]	极低	NR	低	NR	NR	C
Zhong & Liu's Scale[49]	中	高	高	高	NR	B
AI Literacy Tool[26]	高	NR	高	NR	NR	B
AILST[50]	中	高	高	高	NR	B
AILIT-S[27]	NR	高	高	高	NR	B
TAICS[51]	中	高	高	高	NR	B
小学生智能素养量表[52]	中	高	高	NR	NR	B

4 讨论

4.1 人工智能素养量表维度内涵与发展

AI 素养的测评体系正在从基础能力导向向综合素养导向进阶。当前量表在测量知识、技能和伦理等基础维度的同时，正逐步拓展至社会情感和高阶认知等综合维度。值得注意的是，人工智能素养的提升不仅依赖于技术能力，更受心理动机影响[56]。尽管部分量表已纳入对 AI 态度的测量，并触及自我效能概念，但整体上仍缺乏对内在驱动因素的评估。基于此，建议未来研究在量表中增设心理认知维度，以量化个体使用 AI 技术的意愿和信心水平。

相较于信息素养和数字素养，高阶思维是 AI 素养的核心[57]。随着研究的深入，越来越多的量表开始涵盖批判思维和人机协同创新能力等高阶思维维度。然而，目前已开发的部分量表条目设置主要基于早期技术场景，如传统机器学习，更偏重基础技能测评，忽视复杂问题解决与创造性思维。这可能导致量表难以有效捕捉生成式人工智能等新兴技术所要求的高级认知能力。因此，应建立“技术演进－维度迭代”关联机制，定期通过德尔菲法或焦点小组访谈更新条目池[58]。具体而言，在量表开发与修订中应突破基础技能测评局限，重视高阶思维导向，通过增设“复杂问题解决”“生成式创作”等深度创意维度和条目，引导在使用大语言模型工具完成信息检索、文本翻译等基础上，重视深层次思维创新[59]。

4.2 人工智能素养量表方法学质量亟须提升，测量学属性有待完善

量表内容效度直接关系量表质量，被视为最重要的测量学属性，而其开发阶段的方法学更是决定内容效度质量的关键因素。研究发现，虽然大多数量表在开发过程中使用了合适的理论框架确定量表构念属性，并采用德尔菲法分析及确定题项[60]。但现有文献对内容效度的报告仍存在不完整、不规范等问题。具体表现为，21 项涉及内容效度验证的研究，未对内容效度的相关性、全面性、可理解性进行详细描述，且普遍存在缺乏明确的访谈提纲及未报告详细的访谈信息等问题。还应注意的是，偏倚风险清单中强调测试对象与专家评价具有同等重要性，但只有 7 项研究[25,33－34,41,43,46,52]进行了预实验，两项研究[30,45]对测试对象进行了认知访谈。

现有量表测量学属性验证仍不够全面。具体而言，尽管 23 项研究评价了量表结构效度，但其中 5 项[33,37,39,42,52]未报告模型拟合优度指标。在 25 项计算内部一致性的研究中，有 6 项[25,28,32,43,49,52]未说明各维度 Cronbach’s α 值。此外，部分量表虽然在结构效度和内部一致性方面表现良好，但未对跨文化效度、稳定性、假设检验、反应度等测量学属性进行验证。值得关注的是，越来越多的研究开始采用已开发的量表进行不同人群的 AI 素养分析，例如 Moodi G A 等[61]在伊朗医学生群体中验证了波斯语版 MAIRS－MS 量表，Ziapour A 等[62]使用 MAIRS－MS 分析医学生 AI 素养的影响因素。这表明对成熟量表的跨文化效度验证以及影响因素研究正不断扩展。因此，建议未来研究应遵循 COSMIN 指南规范研究设计，并对量表的测量学属性进行更新和验证。

4.3 针对不同人群的人工智能素养量表选择与推荐

由于量表质量证据不同，本研究主要基于内容效度、结构效度、内部一致性 3 方面，针对不同人群形成推荐。对于一般人群，AILS[31]在开发过程中经过专家小组资格审核，并且有高质量证据支持结构效度与内部一致性，整体可靠性较高。针对小学生群体，Chung K 等[26]开发的 AI Literacy Tool 经过两轮德尔菲专家验证并达成共识，可较好地保证内容效度。面向中学生群体，推荐使用高中生智能素养调查问卷[33]和 AILQ[43]，二者不仅在结构效度和内部一致性方面表现良好，且均基于测试对象进行了预实验。对于大学生而言，AILIT[25]和 ChatGPT Literacy Scale[45]是较优选择，其内容效度的评价同时整合了专家和测试对象双方意见，且结构效度和内部一致性方面均显示出高质量证据。就具体专业而言，针对医学生首选 MAIRS－MS[30]，针对师范生推荐师范生智能素养自评工具[29]。由于目前未有被评为 A 级的量表，因此上述量表可暂时作为推

荐，其测量学属性仍待检验。

上述针对不同教育阶段和特定人群的 AI 素养测评工具的应用范围仍存在局限。具体而言，虽然已从学生群体和教育培训领域拓展至非专业人群，并依据不同学习阶段开发相应测评工具，但尚未出现针对不同职业群体的 AI 素养量表。此外，即便量表区分了教育阶段（如中学生、大学生），仍然缺乏对学科或专业差异的进一步细化。因此，未来研究应拓展量表开发的应用场景，结合不同学科特点构建差异化框架，以开发更具适配性的测评工具，助力各领域 AI 素养水平的精准评估与提升[63]。

5 结语

本研究系统综述了国内外人工智能素养测评工具的基本特征、维度内涵、方法学质量与测量学属性，梳理了该类工具的发展脉络与共性特征，并指出现有工具在实际应用中的不足。尽管部分量表显示出良好的适用性与推广潜力，但仍存在方法学质量不足、测量学属性验证不全面等问题。

受文献纳入与排除标准限制，本研究基于 COSMIN 指南所纳入的研究不包括会议论文、预印本及除中英文外的其他语种文献，也未涵盖仅提出指标框架而未完成量表构建或测量学验证的研究，这可能对结果的全面性产生一定影响。未来研究应进一步聚焦 AI 素养的核心内涵，引入项目反应理论方法，优化量表维度设置与动态适应能力，并重点开发面向医务人员、医学生及患者等特定群体的专业化测评工具。建议严格遵循 COSMIN 指南规范开发流程，系统验证测量学属性，通过跨文化调试、长期跟踪和临床实用性研究，提升量表的科学性、可靠性和人群针对性，推动 AI 素养测评工具在医学领域的全面发展与实践应用。

作者贡献：林燕负责数据收集与分析、论文撰写与修订；孙海霞负责研究设计、论文修订；蒲思樾负责数据收集与分析、论文修订；陈振丽负责提供指导、论文修订；钱庆负责提供指导、论文审核与修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 UNESCO. AI and education: guidance for policy - makers [EB/OL]. [2025 - 05 - 30]. https://unesdoc.unesco.org/ark:/48223/pf0000376709.

2 赵福君，代洋磊，许静静. 中学生人工智能素养评价指标构建 [J]. 兵团教育学院学报，2023 (6)：68 - 74.

3 ZURKOWSKI P G. The information service environment relationships and priorities. Related paper No. 5 [R]. Washington DC: National Commission on Libraries and Information Science, 1974.

4 GILSTER P. Digital literacy [M]. New York: Wiley, 1997.

5 BURGSTEINER H, KANDLHOFER M, STEINBAUER G. Irobot: teaching the basics of artificial intelligence in high schools [C]. Phoenix: The AAAI Conference on Artificial Intelligence, 2016.

6 郭亚军，寇旭颖，冯思倩，等. 人工智能素养：内涵剖析与评估标准构建 [J]. 图书馆论坛，2025，45 (2)：42 - 50.

7 尹开国. 人工智能素养：提出背景、概念界定与构成要素 [J]. 图书与情报，2024 (3)：60 - 68.

8 PORIA S, CAMBRIA E, BAJPAI R, et al. A review of affective computing: from unimodal analysis to multimodal fusion [J]. Information fusion, 2017, 37 (9): 98 - 125.

9 ACRL. Information literacy competency standards for higher education [EB/OL]. [2025 - 08 - 16]. http://www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm.

10 UNESCO. A global framework of reference on digital literacy skills for indicator 4.4.2 [EB/OL]. [2025 - 08 - 16]. https://uis.unesco.org/sites/default/files/documents/ip51 - global - framework - reference - digital - literacy - skills - 2018 - en.pdf.

11 UNESCO. AI competency framework for students [EB/OL]. [2025 - 08 - 16]. https://unesdoc.unesco.org/ark:/48223/pf0000391105.

12 石映辉，彭常玲，吴砥，等. 中小学生信息素养评价指标体系研究 [J]. 中国电化教育，2018 (8)：73 - 77，93.

13 陈明明，陈雨. 中国居民数字素养的基本内涵、水平测度及结构特征 [J]. 电子政务，2024 (9)：91 - 101.

14 耿瑞利，孙瑜. 面向战略需求的数字素养：概念内涵、框架体系与测评指标 [J]. 图书馆理论与实践，2024 (2)：98 - 106.

15 施雨，茆意宏. 人工智能素养的概念、框架与教育

[J]. 图书馆论坛, 2024, 44 (11): 90 - 100.

16 LONG D, MEGERKO B. What is AI literacy? Competencies and design considerations [C]. Honolulu: The 2020 CHI Conference on Human Factors in Computing Systems, 2020.

17 张银荣, 杨刚, 徐佳艳, 等. 人工智能素养模型构建及其实施路径 [J]. 现代教育技术, 2022 (3): 42 - 50.

18 郑勤华, 覃梦媛, 李爽. 人机协同时代智能素养的理论模型研究 [J]. 复旦教育论坛, 2021 (1): 52 - 59.

19 KIM S, JANG Y, KIM W, et al. Why and what to teach: AI curriculum for elementary school [C]. Online: The AAAI Conference on Artificial Intelligence, 2021.

20 Digital Promise. AI literacy: a framework to understand, evaluate, and use emerging technology [EB/OL]. [2025 - 08 - 16]. https://hdl.handle.net/20.500.12265/218.

21 OECD, European Commission. AI Literacy Framework [EB/OL]. [2025 - 08 - 16]. https://ailiteracyframework.org/wp - content/uploads/2025/05/AILitFramework_ReviewDraft.pdf.

22 MOKKINK L B, ELSMAN E B M, TERWEE C B. COSMIN guideline for systematic reviews of patient - reported outcome measures version 2.0 [J]. Quality of life research, 2024 (11): 2929 - 2939.

23 GAGNER J J, LAI J Y, MOKKINK L B, et al. COSMIN reporting guideline for studies on measurement properties of patient - reported outcome measures [J]. Quality of life research, 2021 (8): 2197 - 2218.

24 陈祎婷, 沈蓝君, 彭健, 等. 改良版定量系统评价证据分级方法对患者报告结局测量工具的评价 [J]. 解放军护理杂志, 2020 (10): 57 - 60.

25 HORNBERGER M, BEWERSDORFF A, NERDEL C. What do university students know about artificial intelligence? Development and validation of an AI literacy test [J]. Computers and education: artificial intelligence, 2023, 5 (1): 100165.

26 CHUNG K, KIM S, JANG Y, et al. Developing an AI literacy diagnostic tool for elementary school students [J]. Education and information technologies, 2025, 30 (11): 1013 - 1044.

27 HORNBERGER M, BEWERSDORFF A, SCHIFF D S, et al. Development and validation of a short AI literacy test (AILIT - S) for university students [J]. Computers in human behavior: artificial humans, 2025, 5 (11): 100176.

28 TENBERGA I, DANIELA L. Artificial intelligence literacy competencies for teachers through self - assessment tools [J]. Sustainability, 2024, 16 (23): 10386.

29 王欢. 师范生人工智能素养自评工具开发研究 [D]. 贵阳: 贵州师范大学, 2021.

30 KARACA O, ÇALIŞKAN S A, DEMIR K. Medical artificial intelligence readiness scale for medical students (MAIRS - MS): development, validity and reliability study [J]. BMC medical education, 2021, 21 (1): 112.

31 WANG B, RAU P L P, YUAN T. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale [J]. Behaviour & information technology, 2022 (9): 1324 - 1337.

32 章恒远. 中小学生人工智能素养测评工具研究 [D]. 上海: 华东师范大学, 2023.

33 周澍云. 高中生人工智能素养评价指标体系构建研究 [D]. 上海: 华东师范大学, 2023.

34 李佳. 面向大学生的人工智能素养评价工具开发研究 [D]. 北京: 北京邮电大学, 2023.

35 CAROLUS A, KOCH J M, STRAKA S, et al. MAILS - Meta AI literacy scale: development and testing of an AI literacy questionnaire based on well - founded competency models and psychological change - and meta - competencies [J]. Computers in human behavior: artificial humans, 2023, 1 (2): 100014.

36 HWANG H S, ZHU L C, CUI Q. Development and validation of a digital literacy scale in the artificial intelligence era for college students [J]. KSII transactions on internet and information systems, 2023 (8): 2241 - 2258.

37 LAUPICHLER M C, ASTER A, HAVERKAMP N, et al. Development of the "Scale for the assessment of non - experts' AI literacy" - an exploratory factor analysis [J]. Computers in human behavior reports, 2023, 12 (1): 100338.

38 LAUPICHLER M C, ASTER A, PERSCHEWSKI J O, et al. Evaluating AI courses: a valid and reliable instrument for assessing artificial - intelligence learning through comparative self - assessment [J]. Education sciences, 2023 (10): 1 - 12.

39 CHAN C K Y, ZHOU W. An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI [J]. Smart learning environments, 2023 (10): 1 - 22.

40 WANG Y Y, CHUANG Y W. Artificial intelligence self - efficacy: scale development and validation [J]. Education and information technologies, 2024 (29): 4785 - 4808.

41 杨欢. 公务员智能素养的结构维度和量表开发 [D].

贵阳：贵州大学，2024.

42 冯剑峰，姜浩哲，刘珈宏．面向人机协同的教师数智素养：测评框架、现状审视与优化路径［J］．教育发展研究，2024（10）：21－29.

43 NG D T K, WU W, LEUNG J K L, et al. Design and validation of the AI literacy questionnaire: the affective, behavioural, cognitive and ethical approach［J］. British journal of educational technology, 2024（3）: 1082－1104.

44 苏文成，郭浩然，卢章平，等．我国高校学生群体人工智能素养评价指标体系构建及实效性验证［J］．图书馆建设，2024（2）：81－93.

45 LEE S, PARK G. Development and validation of ChatGPT literacy scale［J］. Current psychology, 2024（43）: 18992－19004.

46 MA S, CHEN Z Z. The development and validation of the artificial intelligence literacy scale for Chinese college students（AILS－CCS）［J］. IEEE access, 2024（12）: 146419.

47 NONG, Y, CUI, J, HE, Y, et al. Development and validation of an AI literacy scale［J］. Journal of artificial intelligence research, 2024, 1（1）: 17－26.

48 KOCH M J, CAROLUS A, WIENRICH C, et al. Meta AI literacy scale: further validation and development of a short version［J］. Heliyon, 2024（21）: e39686.

49 ZHONG B C, LIU X F. Evaluating AI literacy of secondary students: framework and scale development［J］. Computers & education, 2025（227）: 105230.

50 NING Y, ZHANG W, YAO D, et al. Development and validation of the artificial intelligence literacy scale for teachers（AILST）［J］. Education and information technologies, 2025, 30（3）: 17769－17803.

51 CHIU T K F, AHMAD Z, ÇOBAN M. Development and validation of teacher artificial intelligence（AI）competence self－efficacy（TAICS）scale［J］. Education and information technologies, 2025, 30（10）: 6667 - 6685.

52 崔容华．小学中高年级学生人工智能素养测评工具开发及影响因素研究［D］．上海：上海师范大学，2025.

53 WEBER P, PINSKI M, BAUM L. Toward an objective measurement of AI literacy［C］. Nanchang: The Pacific Asia Conference on Information Systems（PACIS）, 2023.

54 张晓路，李睿，彭馨，等．基于 COSMIN 共识标准的患者报告结局测量理论与概念模型研究［J］．中国卫生经济，2024（10）：1－15.

55 GUYATT G H, OXMAN A D, VIST G E, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations［J］. BMJ, 2008（336）: 924－926.

56 SCHERER R, SIDDIQ F, TONDEUR J. The technology acceptance model（TAM）: a meta－analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education［J］. Computers & education, 2019（128）: 13－35.

57 NG D T K, LEUNG J K L, CHU K W S, et al. AI literacy: definition, teaching, evaluation and ethical issues［C］. Salt Lake City: The Association for Information Science and Technology, 2021.

58 NIEDERBERGER M, KÖBERICH S. Coming to consensus: the Delphi technique［J］. European journal of cardiovascular nursing, 2021（7）: 692－695.

59 SONDERGELD T A, JOHNSON C C. Development and validation of a 21st century skills assessment: using an iterative multimethod approach［J］. School science and mathematics, 2019（6）: 312－326.

60 张亦然，王康美，朱盛财，等．中文版高血压患者自我管理评估工具的系统评价：基于 COSMIN 指南［J］．护理学报，2024（19）：52－57.

61 MOODI G A, MOGHADASIN M, EMADZADEH A, et al. Psychometric properties of the persian version of the medical artificial intelligence readiness scale for medical students（MAIRS－MS）［J］. BMC medical education, 2023（1）: 577.

62 ZIAPOUR A, DARABI F, JANJANI P, et al. Factors affecting medical artificial intelligence（AI）readiness among medical students: taking stock and looking forward［J］. BMC medical education, 2025（1）: 264.

63 CHEE H, AHN S, LEE J. A competency framework for AI literacy: variations by different learner groups and an implied learning pathway［J］. British journal of educational technology, 2024（3）: 1－37.