



Development and Psychometric Validation of the Artificial Intelligence Literacy Scale for University Students: Evidence from Exploratory and Confirmatory Factor Analysis

Aras Bozkurt^{1*}

Anadolu University, Türkiye

Chadia Haddad²

Institut National de Santé Publique, d'Épidémiologie
Clinique et de Toxicologie-Liban, Dekwaneh, Lebanon

Article Info

Article history:

Received: March 21, 2026

Revised: April 25, 2026

Accepted: June 25, 2026

Keywords:

Artificial intelligence literacy;
Confirmatory factor analysis;
Exploratory factor analysis;
Higher education;
Psychometric validation.

Abstract

Background of study: Artificial intelligence (AI) is now embedded in university learning, assessment, research, and academic decision-making, increasing the need for psychometrically sound measures of students' AI literacy. Existing instruments differ substantially in scope, dimensionality, and evidence of validity. Aims and scope of paper: This study develops the Artificial Intelligence Literacy Scale for University Students (AILS-U), a concise multidimensional self-report instrument designed for cross-disciplinary higher-education contexts. Methods: Following best-practice scale-development procedures, an initial item pool was subjected to expert content validation, cognitive review, pilot testing, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA). Because no empirical dataset was supplied for this manuscript draft, numerical results are transparently demonstrated using a synthetic sample ($N = 742$; EFA $n = 371$; CFA $n = 371$) and must be replaced before submission. Result: The illustrative EFA supports a 24-item, five-factor structure covering conceptual understanding, applied use and human-AI collaboration, critical evaluation and verification, ethical and responsible use, and human agency and self-regulation. The illustrative CFA shows good fit ($CFI = .958$, $TLI = .952$, $RMSEA = .050$, $SRMR = .045$), with strong reliability and convergent/discriminant validity. Conclusion: The proposed AILS-U offers a theoretically integrated framework for assessing university students' AI literacy, but publication requires validation with verified participant data and cross-context replication.

To cite this article: Bozkurt, A & Haddad C (2026). *Development and Psychometric Validation of the Artificial Intelligence Literacy Scale for University Students: Evidence from Exploratory and Confirmatory Factor Analysis*. Journal of Quantitative Research and Theory, 1(1), 25-35

INTRODUCTION

Background of the study

Artificial intelligence has become a routine component of university study rather than a specialist technology used only by computer scientists. Students increasingly interact with generative AI assistants, recommender systems, automated feedback, translation tools, learning analytics, coding assistants, and AI-enabled research applications. This diffusion expands opportunities for explanation, feedback, ideation, personalization, and productivity, but it also introduces new demands related to factual verification, privacy, bias, attribution, authorship, academic integrity, and the preservation of human judgment. Consequently, higher education requires more than access to AI tools; it requires students who can understand what AI can and cannot do, use it purposefully, critically evaluate its outputs, act ethically, and regulate their own dependence on automated assistance.

AI literacy provides a framework for these capabilities. Long and Magerko (2020) conceptualized AI literacy as competencies that enable people to critically evaluate AI, communicate and collaborate with AI, and use AI effectively across settings. Ng et al. (2021) organized AI literacy around understanding, use, evaluation, creation, and ethical issues. More recent competency frameworks similarly emphasize the combination of knowledge, skills, values, critical judgment, and human

* Corresponding author:

Aras Bozkurt, Anadolu University, Türkiye
arasbozkurt@gmail.com

agency. UNESCO (2024), for example, structures student AI competency around a human-centred mindset, ethics of AI, AI techniques and applications, and AI system design. These perspectives indicate that AI literacy is multidimensional and cannot be represented adequately by frequency of AI use or confidence with digital tools alone.

The rapid growth of generative AI has intensified the measurement problem. Students may appear highly competent because they use AI frequently, yet frequent use can coexist with weak conceptual understanding, poor source verification, inappropriate disclosure of personal or institutional data, or excessive cognitive offloading. Conversely, students with strong critical and ethical judgment may use AI selectively rather than intensively. A valid scale must therefore distinguish between operational fluency and deeper capacities such as evaluation, responsibility, and human agency.

Rationale of the study

A growing set of AI literacy instruments has emerged. Wang, Rau, and Yuan (2023) developed a 12-item Artificial Intelligence Literacy Scale organized around awareness, use, evaluation, and ethics. Carolus et al. (2023) proposed the Meta AI Literacy Scale (MAILS), which measures several AI-literacy facets together with psychological competencies relevant to AI use. Laupichler et al. (2023) developed the Scale for the Assessment of Non-Experts' AI Literacy (SNAIL), whose exploratory factor analysis identified technical understanding, critical appraisal, and practical application. Hornberger, Bewersdorff, and Nerdel (2023) developed an objective AI literacy test for university students, demonstrating the importance of distinguishing knowledge tests from self-report competence measures. Later work provided additional validation and short forms, while discipline-specific instruments have emerged for domains such as design and academic writing.

Despite this progress, the measurement landscape remains fragmented. Lintner's (2024) systematic review identified substantial variation in target populations, construct definitions, and psychometric procedures; many instruments showed acceptable structural validity and internal consistency, but evidence for content validity, cross-cultural validity, reliability over time, responsiveness, and interpretability was often limited. The field has also evolved rapidly since generative AI became common in higher education. Recent instruments increasingly incorporate practical application and evaluation, yet the ability to maintain human agency and self-regulate AI-supported learning remains less consistently represented as a distinct student competency.

The present study therefore develops the Artificial Intelligence Literacy Scale for University Students (AILS-U). The scale is intended as a concise, cross-disciplinary self-report instrument rather than an objective knowledge examination. Its conceptual model integrates established AI-literacy dimensions with a distinct human-agency and self-regulation dimension. This addition reflects a central challenge of university AI use: students must know not only how to obtain useful AI outputs, but also when to question, limit, disclose, or avoid AI assistance so that learning, authorship, and responsibility remain meaningfully human-led.

Literature review

Conceptual understanding represents students' grasp of foundational AI ideas and limitations. A literate student should recognize that many AI systems identify statistical patterns rather than reason or understand in the same way humans do, that outputs depend on training data and system design, and that generative outputs can be plausible without being true. Conceptual knowledge supports realistic expectations and reduces anthropomorphic or deterministic interpretations of AI. Objective testing research among university students confirms meaningful variation in foundational AI knowledge and suggests that technical background and prior AI experience can influence performance (Hornberger et al., 2023).

Applied use and human-AI collaboration capture the practical ability to select, prompt, refine, and integrate AI tools into authentic academic workflows. This dimension goes beyond exposure or frequency of use by emphasizing purposeful interaction. Students need to translate learning goals into appropriate AI-supported tasks, refine prompts or instructions, combine AI output with disciplinary knowledge, and iteratively improve results. The collaborative emphasis is consistent with Long and Magerko's (2020) view that AI literacy includes effective communication and collaboration with AI and with UNESCO's (2024) progression from understanding to application and creation.

Critical evaluation and verification address the epistemic risks of AI-generated information. Generative systems may produce fabricated citations, outdated information, biased representations, confident errors, or contextually inappropriate recommendations. AI-literate students therefore need to verify factual claims, compare outputs with credible sources, evaluate uncertainty, and recognize when a task requires expert or primary-source judgment. This dimension overlaps with the critical appraisal factors identified in SNAIL and with evaluation dimensions in Wang et al. (2023), but it is operationalized here explicitly around university learning and scholarly information practices.

Ethical and responsible use concerns privacy, fairness, intellectual property, transparency, academic integrity, accountability, and the consequences of AI-supported decisions. Ethical literacy is not an optional add-on because the legitimacy of AI use depends partly on context. The same tool may be appropriate for formative feedback but prohibited in a closed assessment; a student may use AI for language support while still being responsible for source verification and disclosure. Ng et al. (2021), Wang et al. (2023), UNESCO (2024), and recent scale-development studies consistently identify ethics as a core component of AI literacy.

Human agency and self-regulation refer to the student's ability to remain the primary decision maker in AI-supported learning. This includes recognizing when AI use may weaken independent thinking, choosing when not to use AI, preserving ownership of reasoning and writing, monitoring dependence, and reflecting on whether AI is actually improving learning. The dimension is consistent with UNESCO's human-centred orientation and responds to the expanding literature on over-reliance and automation dependence. It also differentiates AI literacy from narrow operational skill by making purposeful non-use a possible indicator of competence.

Gap analysis

Four gaps motivate the AILS-U. First, existing scales differ substantially in dimensional structure, ranging from three-factor to multidimensional models, and there is still no stable consensus regarding which dimensions are essential for university students. Second, many measures were developed before or during the early diffusion of generative AI and therefore give limited attention to self-regulation and human agency in AI-supported learning. Third, some instruments are designed for general adults, specific disciplines, or objective knowledge assessment, making direct institutional use across diverse university programs difficult. Fourth, psychometric reviews continue to call for stronger content-validation procedures, independent exploratory and confirmatory samples, discriminant-validity evidence, and transparent reporting of item retention and factor decisions (Lintner, 2024). The AILS-U addresses these gaps through a concise university-focused model and a staged development process using expert review, cognitive refinement, EFA, and CFA.

Purpose or research questions of the study

The purpose of this study is to develop and psychometrically validate a multidimensional Artificial Intelligence Literacy Scale for University Students (AILS-U). The study is guided by four research questions: (RQ1) What latent factor structure emerges from the proposed AI-literacy item pool? (RQ2) Does the emergent structure demonstrate acceptable model fit in an independent confirmatory sample? (RQ3) Do the retained dimensions demonstrate adequate internal consistency, convergent validity, and discriminant validity? (RQ4) Can the resulting instrument provide an interpretable and practically usable profile of university students' conceptual, applied, critical, ethical, and self-regulatory AI literacy?

Table 1. Proposed conceptual domains of the AILS-U

Domain	Definition	Primary theoretical basis
Conceptual Understanding	Knowledge of foundational AI concepts, capabilities, data dependence, uncertainty, and limitations.	Long & Magerko (2020); Ng et al. (2021); UNESCO (2024)
Applied Use & Human-AI Collaboration	Purposeful selection, prompting, iterative interaction, and integration of AI into academic tasks.	Long & Magerko (2020); Carolus et al. (2023); UNESCO (2024)
Critical Evaluation & Verification	Ability to assess accuracy, bias, uncertainty, evidence quality, and the need for verification.	Wang et al. (2023); Laupichler et al. (2023)

Domain	Definition	Primary theoretical basis
Ethical & Responsible Use	Privacy, fairness, disclosure, intellectual property, academic integrity, and accountability.	Ng et al. (2021); Wang et al. (2023); UNESCO (2024)
Human Agency & Self-Regulation	Ability to regulate reliance on AI, preserve independent reasoning, and decide when AI should or should not be used.	UNESCO (2024); human-centred AI literacy perspective

METHOD

Research Design

The study follows a sequential quantitative scale-development design informed by established psychometric guidance (Worthington & Whittaker, 2006; Boateng et al., 2018; DeVellis & Thorpe, 2021). The planned process includes construct definition, deductive item generation, expert content validation, cognitive review, pilot testing, item analysis, exploratory factor analysis, confirmatory factor analysis, and reliability/validity evaluation. The scale is conceptualized as a reflective multidimensional self-report measure. Each dimension represents a latent capability orientation inferred from multiple observed items.

Important manuscript-development notice: the user did not provide an empirical dataset for the present drafting task. Therefore, all participant counts, factor loadings, fit indices, reliability coefficients, correlations, and group comparisons reported below are synthetic demonstration values created solely to show how a complete JQRT article can be structured. They are not empirical findings and must be replaced with results from verified data before scholarly submission or publication.

Participant

For empirical implementation, the target population is undergraduate and postgraduate students who have used at least one AI-enabled tool for academic purposes during the previous twelve months. A multi-institutional sample is recommended so that the instrument reflects disciplinary and institutional diversity rather than the practices of a single campus. Eligibility should be confirmed through a screening item, and the questionnaire should include students from social sciences and humanities, business and economics, science and engineering, health-related fields, and other programs.

The illustrative psychometric reporting uses a synthetic total sample of 742 students from six hypothetical higher-education institutions. The dataset is divided randomly into two independent subsamples of equal size: Sample A ($n = 371$) for EFA and Sample B ($n = 371$) for CFA. In the synthetic profile, 72.1% are undergraduate students and 27.9% are postgraduate students; 51.6% identify as women, 47.4% as men, and 1.0% select another/prefer not to say category. Approximately 39% report prior formal AI-related coursework or structured training. These proportions are illustrative only.

Population and the methods of sampling / Instrumentation

A stratified or proportionate sampling strategy across faculties and study levels is preferred. Where probability sampling is not feasible, quota-based purposive recruitment should be reported transparently, including institutional participation, response rate, exclusion criteria, and missing-data patterns. Scale-development guidance recommends samples sufficiently large for stable factor extraction and independent confirmation; the use of separate EFA and CFA samples reduces capitalization on chance and provides a stronger test of structural replicability (Worthington & Whittaker, 2006; Boateng et al., 2018).

Item generation began conceptually with 45 statements derived from AI-literacy frameworks and existing instruments, with deliberate avoidance of verbatim replication of proprietary or previously published scales. Eight hypothetical experts in educational technology, psychometrics, AI ethics, higher education, and information literacy independently rated relevance, clarity, and representativeness. Following Lynn's (1986) content-validity logic, items with weak relevance or redundancy were revised or removed. The illustrative content-validation process retained 35 items with an average scale-level content validity index (S-CVI/Ave) of .92. Cognitive review with a small group of students was then used to identify ambiguous language, double-barreled wording, discipline-specific jargon, and items that could be interpreted as measuring general digital

confidence rather than AI literacy. A pilot administration reduced the working set to 30 items before formal factor analysis.

Instrument

The AILS-U uses a five-point Likert-type response scale ranging from 1 = strongly disagree to 5 = strongly agree. Higher scores indicate stronger self-perceived AI literacy. Items are written to assess transferable university-level competence rather than mastery of a particular AI product. The final illustrative 24-item version contains five dimensions: Conceptual Understanding (5 items), Applied Use and Human-AI Collaboration (5 items), Critical Evaluation and Verification (5 items), Ethical and Responsible Use (5 items), and Human Agency and Self-Regulation (4 items). Dimension means can be reported separately; an overall mean may be reported only when higher-order structure is supported and should not obscure meaningful profile differences.

Table 2. Final illustrative AILS-U items

Code	Domain	Item wording
CU1	Conceptual Understanding	I can explain in general terms how many AI systems learn patterns from data.
CU2	Conceptual Understanding	I understand that AI-generated answers can be plausible even when they are incorrect.
CU3	Conceptual Understanding	I can distinguish AI-driven systems from simple rule-based automation in common situations.
CU4	Conceptual Understanding	I understand how training data and model design can influence AI outputs.
CU5	Conceptual Understanding	I can describe important limitations of current AI systems without assuming they think like humans.
AU1	Applied Use & Human-AI Collaboration	I can select an AI tool that is appropriate for a specific academic task.
AU2	Applied Use & Human-AI Collaboration	I can formulate and refine instructions or prompts to improve an AI response.
AU3	Applied Use & Human-AI Collaboration	I can combine AI-generated material with my own disciplinary knowledge rather than using it unchanged.
AU4	Applied Use & Human-AI Collaboration	I can work iteratively with AI to compare alternatives, revise ideas, or improve a solution.
AU5	Applied Use & Human-AI Collaboration	I can adapt my use of AI when the first response does not meet the learning goal.
CE1	Critical Evaluation & Verification	I verify important factual claims made by AI using credible sources.
CE2	Critical Evaluation & Verification	I can identify signs that an AI answer may contain fabricated, biased, or unsupported information.
CE3	Critical Evaluation & Verification	I compare AI outputs with primary, scholarly, or authoritative sources when accuracy matters.
CE4	Critical Evaluation & Verification	I consider uncertainty and context before accepting an AI recommendation.
CE5	Critical Evaluation & Verification	I can judge when an AI task requires human expertise or independent checking.
ER1	Ethical & Responsible Use	I avoid entering sensitive personal, institutional, or confidential information into AI systems without permission.
ER2	Ethical & Responsible Use	I disclose or cite AI assistance when academic rules or scholarly practice require it.
ER3	Ethical & Responsible Use	I consider fairness and possible bias when AI affects people or academic decisions.
ER4	Ethical & Responsible Use	I respect copyright, authorship, and academic-integrity requirements when using AI-generated material.
ER5	Ethical & Responsible Use	I remain responsible for the accuracy and consequences of work I submit even when AI assisted me.

Code	Domain	Item wording
HA1	Human Agency & Self-Regulation	I can recognize when AI use is replacing rather than supporting my own thinking.
HA2	Human Agency & Self-Regulation	I deliberately choose not to use AI when independent practice is more valuable for learning.
HA3	Human Agency & Self-Regulation	I monitor whether my reliance on AI is increasing over time.
HA4	Human Agency & Self-Regulation	I adjust how I use AI when it reduces my understanding, effort, or ownership of the work.

Procedures and time frame

An empirical study should obtain ethics approval before recruitment. Participants should receive an information sheet describing the purpose of the study, voluntary participation, expected completion time, data protection procedures, and the right to withdraw before submission. Data collection can be conducted online over approximately six to eight weeks. To reduce careless responding, the survey can include an attention check, minimum completion-time rules, and pre-specified criteria for missingness and patterned responses. For cross-language use, translation and cultural adaptation should follow forward-backward translation and cognitive debriefing rather than simple literal translation.

Analysis plan

The psychometric analysis is conducted in two independent stages. In Sample A, item distributions, missingness, item-total correlations, factorability, and redundancy are examined before EFA. Because Likert responses are ordinal, a polychoric correlation matrix and an oblique factor solution are preferred. The number of factors is determined using parallel analysis, scree inspection, theoretical interpretability, and eigenvalue information rather than the eigenvalue-greater-than-one rule alone. Items are considered for removal when they show low communalities, primary loadings below approximately .50, substantial cross-loadings, or conceptual redundancy. Correlated factors are expected because AI-literacy competencies are related but distinguishable.

In Sample B, CFA tests the factor structure emerging from EFA using an estimator appropriate for ordered categorical indicators, such as robust weighted least squares. Model evaluation reports chi-square, CFI, TLI, RMSEA with confidence intervals, and SRMR, while avoiding rigid reliance on a single cutoff (Hu & Bentler, 1999; Kline, 2023). Internal consistency is examined using Cronbach's alpha and McDonald-type composite reliability estimates; convergent validity is assessed through standardized loadings and average variance extracted (AVE), while discriminant validity is evaluated using factor correlations, the Fornell-Larcker logic, and HTMT ratios (Fornell & Larcker, 1981; Henseler et al., 2015). Competing one-factor and higher-order models are compared to determine whether the scale should be interpreted primarily as a multidimensional profile or as a single general score.

RESULTS AND DISCUSSION

Results

All numerical results in this section are synthetic demonstration values for manuscript development and must be replaced with verified empirical outputs before submission.

Table 3. Illustrative sample profile (synthetic values)

Characteristic	n	%
Total sample	742	100.0
Sample A - EFA	371	50.0
Sample B - CFA	371	50.0
Undergraduate	535	72.1
Postgraduate	207	27.9
Social sciences & humanities	226	30.5
Business & economics	156	21.0
Science & engineering	235	31.7
Health & other fields	125	16.8
Prior formal AI training	289	39.0

Content validation produced high illustrative agreement among the eight experts. Thirty-five of the original 45 items were retained after relevance and clarity review, yielding an S-CVI/Ave of .92. Cognitive review led to wording changes in 11 items, particularly statements that students initially interpreted as measuring general confidence with technology rather than AI-specific literacy. Pilot item analysis removed five additional items because of redundancy, weak corrected item-total correlations, or difficulty differentiating between adjacent conceptual domains. The 30-item form entered EFA.

In Sample A, the Kaiser-Meyer-Olkin measure was .947 and Bartlett's test of sphericity was statistically significant, $\chi^2(435) = 5,632.8$, $p < .001$, supporting factorability. Parallel analysis and the scree plot converged on a five-factor solution. Six items were removed sequentially because of cross-loadings or low communalities. The final 24-item solution explained 68.7% of the common variance. Primary loadings ranged from .61 to .88, all retained communalities exceeded .42, and no retained cross-loading exceeded .30. The five factors were interpretable as Conceptual Understanding, Applied Use and Human-AI Collaboration, Critical Evaluation and Verification, Ethical and Responsible Use, and Human Agency and Self-Regulation.

Table 4. Illustrative exploratory factor analysis results (synthetic values)

Code	Factor	Primary loading	Communality
CU1	Conceptual Understanding	0.74	0.55
CU2	Conceptual Understanding	0.81	0.66
CU3	Conceptual Understanding	0.68	0.49
CU4	Conceptual Understanding	0.77	0.61
CU5	Conceptual Understanding	0.72	0.54
AU1	Applied Use & Collaboration	0.71	0.52
AU2	Applied Use & Collaboration	0.84	0.71
AU3	Applied Use & Collaboration	0.76	0.60
AU4	Applied Use & Collaboration	0.79	0.64
AU5	Applied Use & Collaboration	0.73	0.56
CE1	Critical Evaluation	0.82	0.69
CE2	Critical Evaluation	0.85	0.73
CE3	Critical Evaluation	0.78	0.63
CE4	Critical Evaluation	0.70	0.51
CE5	Critical Evaluation	0.75	0.59
ER1	Ethical & Responsible Use	0.69	0.49
ER2	Ethical & Responsible Use	0.80	0.66
ER3	Ethical & Responsible Use	0.74	0.57
ER4	Ethical & Responsible Use	0.83	0.70
ER5	Ethical & Responsible Use	0.77	0.62
HA1	Human Agency & Self-Regulation	0.76	0.60
HA2	Human Agency & Self-Regulation	0.72	0.55
HA3	Human Agency & Self-Regulation	0.61	0.42
HA4	Human Agency & Self-Regulation	0.88	0.78

CFA in Sample B supported the correlated five-factor model. The illustrative fit statistics were $\chi^2 = 468.2$, $df = 242$, $\chi^2/df = 1.93$, CFI = .958, TLI = .952, RMSEA = .050 (90% CI [.043, .057]), and SRMR = .045. Standardized loadings ranged from .69 to .91. A one-factor model fit substantially worse (CFI = .692, TLI = .661, RMSEA = .132, SRMR = .118), indicating that AI literacy should not be treated as unidimensional. A second-order model with a general AI-literacy factor was acceptable but slightly weaker (CFI = .949, TLI = .943, RMSEA = .057, SRMR = .051), supporting the use of dimension scores as the primary interpretation while permitting a cautious overall score when theoretically justified.

Table 5. Illustrative CFA reliability and validity statistics (synthetic values)

Dimension	Items	Loading range	Cronbach alpha	Omega	CR	AVE
Conceptual Understanding	5	.69-.85	.86	.87	.87	.57
Applied Use & Human-AI Collaboration	5	.71-.88	.89	.90	.90	.64

Dimension	Items	Loading range	Cronbach alpha	Omega	CR	AVE
Critical Evaluation & Verification	5	.73-.91	.91	.92	.92	.70
Ethical & Responsible Use	5	.70-.89	.90	.90	.91	.67
Human Agency & Self-Regulation	4	.69-.87	.85	.86	.86	.61

Reliability and convergent validity were acceptable in the synthetic demonstration. Cronbach's alpha values ranged from .85 to .91, omega estimates from .86 to .92, composite reliability from .86 to .92, and AVE from .57 to .70. HTMT ratios ranged from .49 to .79, remaining below the commonly used .85 criterion and supporting discriminant validity. The strongest factor correlation occurred between Applied Use and Critical Evaluation ($r = .69$), which is theoretically plausible because effective academic use often requires iterative evaluation of AI outputs. The weakest correlation was between Conceptual Understanding and Ethical and Responsible Use ($r = .38$), indicating that technical-conceptual knowledge and ethical judgment are related but clearly distinguishable competencies.

Table 6. Illustrative inter-factor correlations and HTMT ranges (synthetic values)

Factor	CU	AU	CE	ER	HA
CU	1.00	.52	.55	.38	.44
AU	.52	1.00	.69	.50	.58
CE	.55	.69	1.00	.62	.65
ER	.38	.50	.62	1.00	.57
HA	.44	.58	.65	.57	1.00

Note. CU = Conceptual Understanding; AU = Applied Use & Human-AI Collaboration; CE = Critical Evaluation & Verification; ER = Ethical & Responsible Use; HA = Human Agency & Self-Regulation. Illustrative HTMT values ranged from .49 to .79.

Discussion

The proposed AILS-U demonstrates a coherent five-factor structure in the illustrative psychometric workflow. The results suggest that university students' AI literacy is better understood as a profile of related competencies than as a single undifferentiated trait. This conclusion aligns with the broader literature, where AI literacy instruments repeatedly identify separable dimensions involving understanding, use, evaluation, and ethics, even though the number and labeling of factors vary across instruments (Wang et al., 2023; Carolus et al., 2023; Laupichler et al., 2023; Lintner, 2024). The present model preserves those established elements while making human agency and self-regulation explicit.

The Conceptual Understanding factor reflects a foundational layer of literacy. Students need enough knowledge to recognize that AI outputs are shaped by training data, model architecture, probabilistic inference, and system constraints. This does not imply that every student must be able to program or train models. Instead, foundational understanding should support realistic expectations, informed questioning, and resistance to anthropomorphic misconceptions. The distinction between conceptual understanding and practical use also mirrors objective-test research showing that frequent AI exposure does not guarantee accurate knowledge of AI principles (Hornberger et al., 2023).

Applied Use and Human-AI Collaboration emerged as a distinct factor rather than merely a frequency-of-use dimension. This is important for higher education because students increasingly interact with AI through iterative prompting, comparison, revision, and integration into complex tasks. The construct therefore reflects purposeful orchestration rather than passive consumption. Its relationship with Critical Evaluation is expected: students who collaborate effectively with AI must repeatedly judge whether the output is useful, credible, and aligned with disciplinary standards.

Critical Evaluation and Verification appears particularly important in the generative-AI era. The factor emphasizes source checking, bias detection, uncertainty, and recognition of contexts in which human expertise is necessary. Earlier scales already identify evaluation or critical appraisal as a central domain, including Wang et al. (2023) and SNAIL (Laupichler et al., 2023). The AILS-U extends this logic by anchoring evaluation in academic information practices, where fabricated references or inaccurate claims can directly affect scholarly quality.

Ethical and Responsible Use remains a core, separable dimension. The pattern supports frameworks that position ethics alongside knowledge and application rather than treating ethical behavior as an external attitude. University students face context-specific decisions about privacy, disclosure, copyright, academic integrity, and responsibility for submitted work. Ethical competence therefore requires practical judgment about institutional rules and consequences, not simply agreement with abstract principles. This interpretation is consistent with Ng et al. (2021), Wang et al. (2023), and UNESCO's (2024) explicit emphasis on safe and responsible AI use.

The most distinctive feature of the AILS-U is Human Agency and Self-Regulation. The factor captures the ability to notice dependence, protect independent reasoning, choose deliberate non-use, and modify AI behavior when learning quality declines. This dimension is timely because a technically proficient student can still be poorly AI-literate if AI use consistently substitutes for thinking, practice, or authorship. The human-centred emphasis is also consistent with UNESCO's (2024) competency framework, which places human agency and accountability at the foundation of student AI competence. Future empirical research should determine whether this factor replicates across national and disciplinary contexts and whether it predicts learning outcomes beyond general self-regulated learning.

The proposed scale also responds to recent psychometric developments. Lintner (2024) noted that many AI-literacy scales have strong internal consistency and structural validity but more limited evidence for content validity, cross-cultural validation, measurement error, and responsiveness. Recent higher-education scale-development work continues to refine dimensional models, including a four-dimensional 17-item scale by Kang and Park (2026) and discipline-specific validation among design students by Lin and Tan (2026). The AILS-U should therefore be viewed as a theoretically motivated contribution to an evolving measurement ecosystem rather than a replacement for all existing instruments.

Implications

For universities, a validated AILS-U could support needs assessment, curriculum design, orientation programs, AI-literacy interventions, and evaluation of institutional AI initiatives. Dimension-level reporting is particularly useful because two students with the same total score may have different needs: one may understand AI concepts but lack ethical confidence, while another may be an effective user who rarely verifies outputs. Institutions can use these profiles to design targeted learning activities rather than treating AI literacy as a binary competent/not-competent status.

For researchers, the scale provides a candidate measurement model that can be tested alongside academic outcomes, digital competence, self-regulated learning, AI self-efficacy, technology acceptance, critical thinking, and responsible-use behavior. It may also support evaluation of educational interventions. However, researchers should avoid interpreting self-report literacy as equivalent to objectively demonstrated competence. When feasible, the AILS-U should be paired with performance-based items, scenario judgments, or short knowledge tests to reduce shared-method bias and to examine calibration between perceived and actual competence.

Limitations

The draft has several important limitations. Most importantly, the numerical results are synthetic and cannot be interpreted as empirical evidence. Actual publication requires ethics approval, real participant recruitment, transparent data screening, and reproducible factor analyses. Second, the AILS-U is a self-report instrument and is therefore vulnerable to overconfidence, social desirability, and limited self-awareness. Third, the proposed items are written for general university contexts; specialized disciplines such as medicine, law, engineering, or design may require additional domain-specific competencies. Fourth, cross-cultural equivalence, test-retest reliability, measurement error, responsiveness to instruction, and predictive validity have not been established. Fifth, AI technologies evolve rapidly, meaning that item wording must remain sufficiently tool-agnostic to avoid becoming obsolete.

Suggestions

Future studies should validate the AILS-U with large multi-institutional samples and preregister item-retention criteria where feasible. Independent replications should test measurement invariance across gender, academic level, discipline, language, and country. Test-retest reliability and longitudinal responsiveness should be examined so that the instrument can be used to evaluate AI-

literacy interventions. Item-response theory or Rasch analysis may provide additional information about item difficulty, discrimination, and score precision. Researchers should also examine convergent and criterion validity against objective AI-literacy tests, logged AI-use behaviors, scenario-based ethical judgments, and verified academic outcomes. Finally, periodic content review is necessary because emerging AI capabilities may change what competent use, evaluation, and disclosure require.

CONCLUSION

This study proposes the Artificial Intelligence Literacy Scale for University Students (AILS-U), a concise multidimensional instrument designed to assess how students understand, use, evaluate, govern, and self-regulate AI in higher education. The conceptual model integrates five domains: Conceptual Understanding, Applied Use and Human-AI Collaboration, Critical Evaluation and Verification, Ethical and Responsible Use, and Human Agency and Self-Regulation. The illustrative EFA and CFA workflow demonstrates how a 24-item structure can be evaluated using independent samples, factor-retention procedures, reliability analysis, and convergent/discriminant validity assessment. The model's central contribution is to treat AI literacy not only as knowledge or tool proficiency, but also as the ability to preserve judgment, responsibility, and learning agency while working with AI. Because the present numerical results are synthetic, the manuscript should be considered a complete research draft and measurement blueprint rather than publishable empirical evidence. Verified data and independent replication are required before the AILS-U can be claimed as a validated instrument.

ACKNOWLEDGMENT (IF ANY)

The authors acknowledge the scholarly contributions of researchers whose AI-literacy frameworks and measurement studies informed the conceptual development of this manuscript. Because the current version is a manuscript-development draft using synthetic demonstration values, no empirical participants or institutions are represented in the reported results. Any future funding source, ethics approval number, participating institutions, expert-panel contributions, or data-support acknowledgments should be added only after the empirical study has been completed and verified.

AUTHOR CONTRIBUTIONS STATEMENT

AA: conceptualization, literature synthesis, construct definition, item development, methodology, psychometric analysis plan, visualization, and writing-original draft. BB: content validation design, methodology review, supervision, interpretation, writing-review and editing, and project administration. Both authors approved the final manuscript and agree to be accountable for the integrity of the work. Replace AA and BB with the authors' actual initials before submission.

REFERENCES

- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). 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. *Computers in Human Behavior: Artificial Humans*, 1(2), 100014. <https://doi.org/10.1016/j.chbah.2023.100014>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.1177/002224378101800104>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43, 115-135. <https://doi.org/10.1007/s11747-014-0403-8>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185. <https://doi.org/10.1007/BF02289447>
- Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5, 100165. <https://doi.org/10.1016/j.caeai.2023.100165>

- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Kang, K. Y., & Park, J.-H. (2026). Development and validation of a multidimensional AI literacy scale for higher education students: A mixed-method study. *Journal of Librarianship and Information Science*. Advance online publication. <https://doi.org/10.1177/09610006261425235>
- Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.
- Koch, M. J., Carolus, A., Wienrich, C., & Latoschik, M. E. (2024). Meta AI literacy scale: Further validation and development of a short version. *Heliyon*, 10(21), e39686. <https://doi.org/10.1016/j.heliyon.2024.e39686>
- Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy" - An exploratory factor analysis. *Computers in Human Behavior Reports*, 12, 100338. <https://doi.org/10.1016/j.chbr.2023.100338>
- Lin, S., & Tan, Z. (2026). AI Literacy Scale for Undergraduate Design Students: Development and validation. *SAGE Open*. Advance online publication. <https://doi.org/10.1177/21582440261463131>
- Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning*, 9, 50. <https://doi.org/10.1038/s41539-024-00264-4>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-385. <https://doi.org/10.1097/00006199-198611000-00017>
- Ma, S., & Chen, Z. (2024). The development and validation of the Artificial Intelligence Literacy Scale for Chinese College Students (AII-S-CCS). *IEEE Access*, 12, 146419-146429. <https://doi.org/10.1109/ACCESS.2024.3468378>
- Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO. <https://doi.org/10.54675/JKJB9835>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324-1337. <https://doi.org/10.1080/0144929X.2022.2072768>
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838. <https://doi.org/10.1177/0011000006288127>