

MEASURING AI LITERACY: A REVIEW OF GENERAL-PURPOSE SCALES AND A SIX-DIMENSIONAL FRAMEWORK

ChunYi Lin

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

Abstract: The rapid development of generative artificial intelligence (AI) has made AI literacy one of the core competencies of citizens in the digital age. Yet current instruments for assessing AI literacy tend to be regionally bounded, lack cross-cultural equivalence, and measure only a narrow range of dimensions, which limits their usefulness for general assessment across cultures and contexts. This paper reviews Chinese- and English-language studies published between 2021 and 2025, combining CiteSpace bibliometric mapping with thematic analysis of 147 screened records, in order to compare the theoretical foundations, dimensional structures, psychometric properties, and cross-cultural adaptability of AI attitude scales (ALS) and generative AI literacy scales (GLAT). The review finds that most existing scales follow a four-dimensional model of cognition–skill–affect–ethics, but rarely address how assessment procedures might be standardized or how cross-cultural adaptation might work in practice. On this basis, the paper proposes a six-dimensional model—cognition, skill, affect, ethics, assessment, and cross-cultural—to provide theoretical support for building a general-purpose AI literacy scale, and it outlines directions and trends for future scale development.

Keywords: AI literacy; AI attitude scale; Generative AI literacy scale; Cross-cultural assessment; Measurement invariance; Literature review

1 INTRODUCTION

Successive waves of AI technology are reshaping how people work and live. From the widespread use of generative AI tools such as ChatGPT to new forms of human–AI collaboration in education, an individual's ability to understand AI, apply it, and judge it ethically—that is, AI literacy—has become a core competency of the digital age [1-2]. Definitions differ in emphasis, but most converge on a combination of knowledge about AI, practical skill in using it, critical evaluation of what it produces, and ethical awareness. Yet even as K-12 and higher-education institutions worldwide have incorporated AI literacy into their curricula, the development of assessment tools has lagged behind. Most existing scales are tied to particular regions or populations (for example, university students or teachers) and have not been validated for cross-cultural equivalence. The cognition–skill–affect–ethics model dominates their dimensional design, while the standardization of assessment procedures, dynamic adaptability, and cultural adaptability receive limited attention [3]. Against this background, two types of instrument—the Generative AI Literacy Scale (GLAT) and the AI Attitude Scale (ALS)—have attracted growing research attention. The former focuses on the critical evaluation of, and co-creation with, generative AI [4]; the latter places greater weight on overall attitudes toward AI and on value judgments about it [5]. Using a literature review method that combines CiteSpace knowledge mapping with thematic analysis, this paper examines authoritative Chinese- and English-language studies from the past five years (2021–2025). Its aim is to identify likely directions and trends in the construction of general-purpose AI literacy scales and to offer a theoretical reference for standardized assessment across cultures and contexts.

Conceptually, AI literacy grew out of earlier work on information and digital literacy, and early studies often treated it as a subset of the latter. That framing could not fully capture the generativity and opacity of AI systems or the complexity of human–AI collaboration. As generative AI spread, researchers came to see the need for a dedicated literacy framework—one that would cover understanding how models work, evaluating their outputs critically, and making and adjusting judgments in ethically difficult situations. This shift echoes the OECD's emphasis on human–AI collaboration and responsible AI [6], and it points the way toward a redesign of scale dimensions.

2 METHOD

This study uses a literature review design that combines bibliometric mapping with thematic analysis. It does not claim the exhaustive coverage of a PRISMA-based systematic review; the aim is to establish how AI literacy is currently measured and where the measurement frameworks fall short.

2.1 Data Sources and Search Strategy

Chinese-language records were retrieved from CNKI and Wanfang Data, and English-language records from Web of Science and Scopus. The search combined the core terms “artificial intelligence literacy”, “AI literacy”, “AI attitude”, and “generative AI competence” with the instrument-related terms “scale”, “questionnaire”, “instrument”, “assessment”, and “measurement”; the corresponding Chinese terms were AI literacy, AI attitudes, scales, and assessment. Only peer-reviewed journal articles published between January 2021 and December 2025 were considered.

2.2 Screening

Records were first screened by title and abstract, and the full texts of the remaining candidates were assessed against three inclusion criteria: the study develops, adapts, or validates an instrument for measuring AI literacy, attitudes toward AI, or generative AI competence; it reports at least one psychometric property of that instrument; and it appears in a peer-reviewed journal. Editorials, opinion pieces, and studies that discuss AI literacy without developing or validating an instrument were excluded. A total of 147 records met the inclusion criteria and were retained for analysis; the sources cited below are those most directly relevant to the argument of this review.

2.3 Analysis

The retained records were analyzed in two passes. CiteSpace was used for keyword co-occurrence and cluster mapping to identify the thematic structure reported in Sections 3 to 5. The records were then coded thematically along four lines: theoretical foundations, dimensional structure, psychometric evidence, and cross-cultural adaptation. What follows is therefore a structured critical synthesis of the field rather than a meta-analytic estimate of effect sizes.

3 THEORETICAL FOUNDATIONS OF EXISTING SCALES: FROM INFORMATION LITERACY TO AI LITERACY

The theoretical basis of AI literacy assessment has shifted from traditional information literacy to AI literacy. Early studies generally treated AI literacy as a subset of digital literacy and retained the classic access–understanding–application framework [7]. An influential early synthesis recast AI literacy as four aspects—knowing and understanding, using and applying, evaluating and creating, and ethical issues—and this reading became the template that most later instruments followed [2]. The European Digital Competence Framework (DigComp 2.2), for instance, mentions algorithmic awareness and the use of AI tools, but it does not provide dimensions tailored to what is distinctive about AI—such as the opacity of generated content or the complexity of human–AI collaboration [7].

With the rise of generative AI, research turned toward more specific theories of AI literacy. The generative AI competence model proposed by Jin et al. divides literacy into four core dimensions—knowledge (understanding how AI works), application (using AI tools), evaluation (judging the reliability of AI outputs), and ethics (recognizing AI-related risks and responsibilities)—and it has served as the theoretical backbone of the GLAT [4]. The model moves beyond the traditional emphasis on technical operation and gives greater weight to critical thinking about AI-generated content and to ethical judgment, aligning closely with the OECD's emphasis on human–AI collaboration and responsible AI [6].

In China, Cai Yingehun and colleagues drew on local educational practice to argue that AI literacy has three components—technical cognition, critical thinking, and ethical awareness—and to stress its continuity with, and evolution from, digital literacy and core competencies [8]. Zhong Baichang's team refined this into three levels: cognitive understanding, applied competence, and ethical judgment. They argue that AI literacy is, at heart, “the capacity of individuals to reconstruct their ways of thinking and behaving in an intelligent age” [9].

The theoretical foundations of the ALS are more eclectic. Beyond digital literacy theory, it draws on the technology acceptance model (TAM), constructivist learning theory, and ethical decision-making theory. TAM, for example, explains how users' perceived usefulness and perceived ease of use shape their willingness to adopt AI tools [10]; constructivist learning theory emphasizes developing AI literacy through collaborative practice in authentic settings [11]; and ethical decision-making theory supplies a principles–consequences–virtues framework for measuring the ethical dimension [12]. Together, these strands have pushed scales away from assessing a single skill and toward capturing a combination of multiple competencies.

Recent work has also paid increasing attention to metacognition and dynamic adaptability. Metacognition concerns how individuals monitor and reflect on their own use of AI; dynamic adaptability concerns the ability to keep updating knowledge and methods in a fast-changing technological environment. Incorporating these elements allows a scale to go beyond static competencies and to capture how learners adjust strategies and transfer what they know to real tasks, which makes assessment more timely and forward-looking.

4 DIMENSIONAL STRUCTURE OF EXISTING SCALES: CONSENSUS AND LIMITS OF THE FOUR-DIMENSIONAL MODEL

Both the GLAT and the ALS are built mainly around the cognition–skill–affect–ethics model, though they differ in what each dimension contains and where they place emphasis.

4.1 Core Components of the Four-Dimensional Model and the Empirical Evidence

International scales such as MAILS (Meta AI Literacy Scale) and AILST (Artificial Intelligence Literacy Scale for Teachers) adopt the same four dimensions: cognition (knowledge of AI and its underlying principles), skill (applying and innovating with AI tools), affect (acceptance of AI and willingness to use it), and ethics (recognizing AI risks and taking responsibility) [13–14]. The same structure recurs in instruments built for other populations: the 32-item AI Literacy Questionnaire (AILQ) confirmed a four-factor affective, behavioral, cognitive, and ethical structure among secondary school students [15]. Other recent instruments organize generative AI competence differently. The Generative AI

Competence Scale (GenAIComp), for instance, is built around five factors—information and data literacy, communication and collaboration, digital content creation, safety and ethics, and problem-solving—rather than the familiar four [16]. Chinese scales, including the Smart Literacy Evaluation System and AILS-CKCE (high school version), follow this framework as well, but they give more weight to social responsibility within a collectivist culture [17-18]. AILS-CKCE, for example, breaks the ethical dimension into privacy protection, fairness and transparency, and social impact—a reflection of the particular attention Chinese education pays to the relationship between technology and society. A related line of domestic work builds models of how university students learn with generative artificial intelligence and derives an evaluation index system from them, which gives the four dimensions a more operational definition [19]. Psychometric evidence generally supports the reliability and validity of the four-dimensional model. The GLAT, which uses a mixed-format assessment (objective items plus situational tasks), reports a Cronbach's α of 0.80 for its cognition and application dimensions, and a confirmatory factor analysis (CFI = 0.93, RMSEA = 0.06) supports the four-factor structure [4]. For the domestic AI-Empowered Foreign Language Learning Literacy Questionnaire (AIELQ) for university students, the affect dimension yields $\alpha = 0.921$ with a test-retest reliability of 0.89, indicating acceptable temporal stability and internal consistency [20]. Comparable evidence has been reported for other recent instruments, among them the ChatGPT Literacy Scale, the AI Literacy Concept Inventory developed for middle-school students, and objective measures designed to offset the weaknesses of self-report formats [21-23]. Table 1 summarizes the representative instruments discussed in this section.

Table 1 Representative Instruments for Measuring AI Literacy and Attitudes toward AI

Scale (author, year)	Target population	Dimensional structure	Measurement format
AILQ (Ng et al., 2024) [15]	Secondary school students	Affective, behavioral, cognitive, ethical	Self-report, 32 items
GenAIComp (Lee et al., 2026) [16]	University students	Information and data literacy; communication and collaboration; content creation; safety and ethics; problem-solving	Self-report, 5 factors
GLAT (Jin et al., 2025) [4]	Post-secondary and general users	Cognition, skill, ethics (three modules)	Mixed: objective items, situational judgment tasks, self-report; 20 items
MAILS (Carolus et al., 2023) [13]	General adult users	Cognition, skill, affect, ethics, plus change and meta-competence	Self-report
AILST (Ning et al., 2025) [14]	Teachers	Cognition, skill, affect, ethics	Self-report
AILS, Turkish version (Yabana Kiremit, 2025) [24]	Healthcare professionals	Cognition, skill, affect, ethics, plus AI-assisted clinical decision-making	Self-report
AIELQ (Zhang, 2025) [20]	University students (foreign language learning)	Cognition, skill, affect, ethics	Self-report
AILS-CKCE, high-school version (Zhu et al., 2023) [18]	Minors (K-12)	Cognition, skill, affect, ethics (privacy protection, fairness and transparency, social impact)	Self-report
Smart Literacy Evaluation System (Li et al., 2025) [17]	University students	Cognition, skill, affect, ethics	Self-report

4.2 Limitations of the Four-Dimensional Model

For all its consensus, the four-dimensional model has clear limitations. First, it reduces the range of dimensions actually assessed. Existing scales concentrate on static competencies such as recalling knowledge or operating tools and give little coverage to dynamic competencies, such as making real-time decisions during human–AI collaboration or responding flexibly to ethical dilemmas [3]. Second, cross-cultural fit is weak. Western scales such as MAILS emphasize individual technical ability and innovation, whereas Eastern scales such as AILS-CKCE emphasize social impact and collective responsibility; transplanting one into the other context can introduce cultural bias. The creative application dimension in Western scales, for instance, encourages users to develop their own AI tools, yet in collectivist cultures a tendency toward risk avoidance may depress scores on that dimension [13]. The problem is not confined to individual instruments: a systematic review of 22 validation studies covering 16 AI literacy scales found that none of them had been tested for cross-cultural validity or for measurement error [25]. Third, assessment procedures are not standardized. Most scales rely on self-report questionnaires, often five-point Likert scales, which are vulnerable to social desirability bias: respondents may overstate their ability in order to appear technologically confident [4].

The four-dimensional model also has blind spots when it comes to problems specific to generative AI, such as hallucination and the amplification of bias. Because generative AI produces highly plausible-looking output, the cognition and skill dimensions alone are poorly suited to detecting potential misinformation. A dedicated assessment and meta-critique dimension is needed to capture how individuals judge the truthfulness and reliability of generated content—one important reason why recent tools such as the GLAT have introduced mixed-format assessment.

5 NEW DIRECTIONS: COMPLEMENTARITY AND INTEGRATION OF THE GLAT AND THE ALS

In response to these limitations, recent research has sought to extend the depth and breadth of assessment by combining the GLAT with the ALS.

5.1 The GLAT: From Self-Report to Mixed-Format Assessment

The GLAT is the first internationally developed mixed-format instrument for generative AI literacy. Its innovation lies in pairing objective tests—situational judgment items and simulated tasks—with a self-report questionnaire. The GLAT developed by Jin et al. contains 20 items covering three modules: basic principles of generative AI (cognition), evaluation of output (skill), and judgment of ethical risk (ethics). Item response theory (IRT) is used to calibrate item difficulty and thereby improve measurement precision [4]. Empirical work shows that the GLAT distinguishes high- and low-literacy groups with good sensitivity (effect size $d = 0.82$), and that its objective task module effectively reduces social desirability bias (correlations fall by 23% relative to a purely self-report questionnaire) [4].

Valuable as the GLAT is, it has two notable weaknesses. The first is its limited cross-cultural equivalence. Its items were designed primarily for English-language contexts, which may disadvantage non-native English speakers—Chinese speakers among them—because languages differ in their logical structure and mode of expression, and respondents may struggle to read the intent of items written for another language context. The second weakness is narrow coverage of the affect and assessment dimensions. The current version of the GLAT does not separately measure the willingness to accept AI or the ability to reflect on AI output, even though both are key to human–AI collaboration. In collaborative settings, a willingness to accept AI shapes the efficiency and quality of the collaboration, while the ability to reflect on AI output helps users make better use of what the system produces and improves outcomes in work and study.

5.2 The ALS: From General to Occupation-Specific

The ALS has moved toward a combination of a general foundation and occupation-specific refinements. Internationally, AILST (teacher version) and AILS (healthcare professional version) adjust the weighting of dimensions to suit their audiences: the teacher version strengthens AI integration in teaching, while the healthcare version adds a subscale on the ethics of AI-assisted clinical decision-making [14, 24]. Domestic scales such as the Smart Literacy Evaluation System, aimed at basic education, emphasize the localization of “AI plus core competencies”—for instance, by incorporating a sense of attachment to family and nation into the ethical dimension [17–18]. The same logic of fitting an instrument to its audience appears in attitude measurement: a validated scale for Chinese university students learning English with generative artificial intelligence adapts the ALS framework to a specific population and subject area [26]. A parallel instrument has been developed for EFL teachers, whose professional context generates a distinct set of measurement demands [27].

The ALS's contribution is that it introduces an assessment dimension. By including items on the critical evaluation of AI-generated content and on reflection about human–AI interaction, it addresses a gap in traditional scales, which measure ability but overlook value judgment. Occupation-specific scales, however, remain at an early stage: most instruments still target general literacy and cannot meet the needs of particular fields such as law or finance [5].

As for integration, some studies have combined the GLAT's objective assessment with the ALS's attitude measure into a two-module instrument that captures both competence and disposition. This approach shows higher ecological validity in higher education and professional training, and it offers a useful model for the modular design of general-purpose scales.

6 CROSS-CULTURAL ADAPTATION AND STANDARDIZATION: KEY BOTTLENECKS IN DEVELOPING A GENERAL-PURPOSE SCALE

Cross-cultural adaptation and standardization are the central bottlenecks in developing a general-purpose AI literacy scale. The challenge is visible both in how different cultural contexts understand the content and value of literacy and in the difficulty of comparing results across studies when assessment procedures lack a common standard.

Existing research points to three ways in which Eastern and Western cultures differ in this respect.

First, the core dimensions carry different weights. Western scales such as MAILS emphasize individual technical ability and innovative application, encouraging users to develop their own AI tools, whereas Eastern scales such as AILS-CKCE emphasize social impact and collective responsibility and attend to how AI affects social equity. Comparisons between Chinese and international scales bear this out: the international Generative AI Competence Scale foregrounds content creation and problem-solving [16], whereas a domestic framework for the digital literacy of international Chinese language teachers stresses evaluation and reflection alongside practice and application [28]. The contrast reflects differing value orientations toward AI literacy in different educational contexts.

Second, ethical standards are interpreted differently across cultures. Privacy protection is treated as an individual right in individualist cultures, whereas collectivist cultures may give priority to the public interest. Such differences mean that a single item can activate different paths of value judgment in different cultural groups, which in turn shapes responses and scores.

Third, the language of assessment carries implicit logical assumptions. In English-language scales, critical thinking is often associated with questioning authority, yet in Chinese the corresponding term may be misread as outright rejection—an ambiguity that localized translation and conceptual clarification can help reduce [29]. The GLAT was developed

mainly for English-language contexts; how well its items and scenarios travel to non-English cultural settings has not been fully tested, leaving open the possibility of cultural bias [4]. The structure of AI literacy also differs markedly across professional backgrounds and occupational groups, and the GLAT has yet to develop mechanisms for adapting to those differences [30].

To address these problems, two recent reviews of these instruments propose a six-dimensional model—cognition, skill, affect, ethics, assessment, and cross-cultural [3, 5]. The cross-cultural dimension requires a scale to include explicit indicators of cultural adaptability, such as recognizing cultural bias and adjusting one's use of AI to fit a local context. A Chinese-language version of the scale, for example, might add an item such as “Would you adjust how much you trust AI-generated content in light of the values of Chinese society?” to measure an individual's capacity for cultural adjustment. This aligns with the view advanced in reviews of generative AI literacy assessment: future work should develop multilingual versions of the GLAT, test their measurement equivalence across cultural settings, adapt culture-specific content, and broaden applicability through cross-cultural research, while deepening understanding of cultural differences in generative AI literacy. Table 2 sets out how the six dimensions differ from the four-dimensional model.

Table 2 The Four-Dimensional Model and the Six-Dimensional Model Compared

Dimension	Focus in the four-dimensional model	Gap identified in this review	What the six-dimensional model adds
Cognition	Knowledge of AI and its underlying principles	Stops at recall; does not reach the opacity of generative models	Understanding of generative mechanisms, including why outputs can be plausible yet wrong
Skill	Applying and innovating with AI tools	Covers tool operation but not real-time decision-making in human–AI collaboration	Process-based assessment of how users adjust and optimize AI-generated content
Affect	Acceptance of AI and willingness to use it	Treated as a single disposition, and rarely linked to actual use	Affect tied to trust calibration and willingness to collaborate with AI systems
Ethics	Recognizing AI risks and taking responsibility	Often reduced to knowledge of ethical rules	Judging and responding to ethical dilemmas in context
Assessment	Not measured separately	Existing scales measure ability but not the capacity to evaluate AI output	Explicit indicators of critical evaluation, including agreement with expert judgment
Cross-cultural	Not measured separately	Measurement invariance is rarely tested	Explicit indicators of cultural adaptability and a multilingual equivalence-testing procedure

The other difficulty in standardization is the lack of a uniform assessment procedure. Existing studies differ in item order, scoring rules, and time limits, which makes results hard to compare across studies. Reviews of generative AI literacy assessment argue for a standard operating procedure modeled on international language tests. Such a procedure would fix rules for sampling from an item bank, standardize time limits and testing conditions, and calibrate parallel versions across languages, thereby reducing the influence of procedural variation on measurement and improving the reproducibility of cross-cultural comparisons. Until progress is made on both cross-cultural equivalence and procedural consistency, the “general” in a general-purpose scale will remain nominal, and its cross-cultural application will be little more than a formality.

7 FUTURE TRENDS AND DESIGN DIRECTIONS: TOWARD A STANDARDIZED, RIGOROUS GENERAL-PURPOSE SCALE

Taken together, the literature suggests the following directions for building a general-purpose AI literacy scale.

7.1 Theory: Integrating Multidimensional Models and Strengthening the Assessment of Dynamic Competencies

Future scales should move beyond the traditional four-dimensional model to include dynamic competencies—decision-making in human–AI collaboration and flexible responses to ethical challenges—along with metacognitive ability, such as reflecting on one's own use of AI. Constructivist learning theory, for example, can inform a situational task module (say, a scenario in which a person and an AI system complete a project together), allowing dynamic literacy to be assessed by observing what respondents actually do.

From a static four-dimensional to a dynamic multidimensional model. Existing scales are mostly built on the cognition–skill–affect–ethics model, but the evidence indicates that the GLAT does not capture the dynamic assessment capabilities specific to generative AI and cannot fully represent users' deeper interaction with AI systems [3]. While retaining the four core dimensions, future frameworks should incorporate dynamic competencies—for example, process-based assessment of how users adjust and optimize AI-generated content across different settings—and metacognitive ability, understood as self-reflection on and monitoring of one's own use of AI. They might also draw on the intelligent-TPACK framework, which extends the intersection of technological, pedagogical, and content knowledge with the knowledge teachers need in order to assess AI-based decisions ethically [31]. Such a multidimensional approach can capture static knowledge and skills while also describing learners' strategic transfer and real-time decision-making in authentic tasks.

7.2 Method: Combining Self-Assessment with Other-Assessment and Promoting Standardized Procedures

Self-report questionnaires alone are easily distorted by bias, so they should be combined with other-assessment—for instance, a teacher's or peer's observation of how someone uses AI—and with objective tests such as a practical examination of AI tool use. One way to improve the GLAT would be to add a supplementary indicator measuring the agreement between a respondent's evaluation of AI-generated content and that of experts, thereby increasing objectivity. Mixed assessment and methodological integration. The strength of the GLAT lies in its mixed format—objective items plus situational tasks—which effectively counters the susceptibility of self-report scales to social desirability bias [4]. Future work should further integrate self-assessment, other-assessment (such as a teacher's or peer's observation of AI use), and behavioral observation into a triangulated measurement system. The literature also recommends developing supplementary instruments for the affect and attitude dimensions to compensate for the GLAT's weakness there [4]. One option is a seven-point Likert scale, which increases response granularity and psychometric sensitivity [5], supported by item analysis and exploratory and confirmatory factor analysis to ensure structural robustness.

7.3 Culture: Establishing a Mechanism for Verifying Cross-Cultural Equivalence

Scale development should follow a standardized translation–back-translation–cultural-adaptation procedure and verify measurement invariance using multilingual samples (Chinese and English, for example). Cross-group invariance testing, such as multi-group confirmatory factor analysis, can be used to compare the consistency of factor structures across cultural groups and so ensure global comparability.

Cross-cultural validation and localization. The literature is explicit that the GLAT needs stronger cross-cultural adaptation and validation: multilingual versions, tests of measurement equivalence across cultural settings, and adjustment of culture-specific content [4]. Cross-cultural research would not only broaden the instrument's applicability but also deepen understanding of cultural differences in generative AI literacy. A standardized translation–back-translation–cultural-adaptation procedure should be established, and multi-group confirmatory factor analysis (MGCFA) used to compare the consistency of factor structures across cultural groups so that the scale is fair and comparable worldwide [4]. To address cultural differences in how the ethical dimension is interpreted, items measuring cultural adaptability could be added—for example, identifying cultural bias or adjusting trust in AI-generated content according to local values [5].

7.4 Technology: Scalability and Multimodal Adaptation

Given the rapid growth of multimodal generative AI, scales need to extend their coverage to new technologies, including the ability to interact with complex AI systems [4]. The literature also suggests that future research should examine how assessment results can be turned into concrete teaching strategies and interventions, and how those interventions can be evaluated, so as to close the loop between research and practice [4]. Structurally, a modular approach is promising: a core module covering general dimensions such as understanding AI principles and ethical judgment, and extension modules adding specialized items for particular fields such as AI-assisted medical decision-making or the design of educational applications [5]. Such a design preserves comparability across contexts while meeting occupation-specific needs.

7.5 Advancing Standardization and Institutionalization

Genuine rigor and standardization call for a tiered assessment system: a general-purpose core scale alongside occupation-specific scales, so that different settings can be assessed precisely. Two key dimensions—assessment ability and cultural adaptability—should also be built into the framework systematically. The former centers on the critical evaluation of AI output; the latter on how well an instrument fits different cultural contexts. Stronger ties to educational practice would let the scale serve not only as a measurement tool but also as a basis for curriculum design, teacher training, and policy-making [32], which is how its educational value is maximized.

8 CONCLUSION

Amid digital transformation and the push toward intelligent systems, assessing AI literacy has become increasingly important and is central to building a society in which humans and AI work together. Constructing a rigorous and comprehensive general-purpose scale is essential for assessing AI literacy accurately. Our examination of two representative instruments, the GLAT and the ALS, shows that although existing tools share a degree of consensus on theoretical foundations, dimensional structure, and psychometric properties, they face challenges in cross-cultural adaptability, coverage of dynamic competencies, and standardization of assessment procedures. Their cross-cultural applicability, for instance, remains limited, which restricts their use in multicultural settings, and their ability to track changes in an individual's AI literacy over time is weak, which undermines the completeness of assessment.

The lack of standardized procedures also raises questions about the fairness and comparability of results. Looking ahead, the development of a general-purpose AI literacy scale should take the six-dimensional model—cognition, skill, affect, ethics, assessment, and cross-cultural—as its core framework. On that basis, self-assessment and other-assessment should be combined so that each compensates for the other's weaknesses, and the balance between generality and specificity should be managed so that the scale applies across a wide range of settings while still accommodating the particular needs of different groups. Cross-cultural equivalence should be verified as well, to ensure that results remain consistent and accurate across cultural contexts. Systematic and thorough refinement along these lines would yield a standardized,

rigorous assessment system and provide a precise and reliable tool for cultivating citizens' AI competence in the digital age.

More broadly, a comprehensive upgrade of AI literacy assessment will require simultaneous and sustained effort on three fronts: theoretical depth, methodological innovation, and cultural inclusiveness. Theoretical depth means probing the essential content and internal logic of AI literacy to give the assessment system a firmer foundation. Methodological innovation means continually exploring and applying new assessment techniques and instruments to make evaluation more rigorous and accurate. Cultural inclusiveness means taking full account of differences across cultural contexts during construction so that results are broadly applicable and comparable. Only then can the limitations of existing scales be overcome and a general assessment system be built that is both internationally comparable—able to align with advanced international standards—and rooted in local practice, well suited to local cultural and social needs. Such a system would provide solid support for talent development and for the complex tasks of social governance in an age of human-machine symbiosis, and would help move a collaborative human-AI society toward a better future.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This study was supported by the Fujian Provincial Social Science Fund Project “Research on Strategies for the Coordinated Development of Artificial Intelligence and Talents' Digital Skills in Fujian Province in the Context of New Quality Productive Forces” (Grant No. FJ2025T019).

REFERENCES

- [1] Long D, Magerko B. What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 2020: 1-16.
- [2] Ng D T K, Leung J K L, Chu S K W, et al. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2021, 2: 100041.
- [3] Lin J, Gong X, Wu Z. A review of research on the Generative AI Literacy Scale (GLAT). *Chinese Journal of Science and Technology*, 2025, 1(4): 273-277.
- [4] Jin Y, Martinez-Maldonado R, Gasevic D, et al. GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, 2025, 9: 100436.
- [5] Lin J, Hu X, Wu Z. A review of research on the Artificial Intelligence Attitude Scale (ALS). *Chinese Journal of Science and Technology*, 2025, 1(4): 267-272.
- [6] OECD. *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. Paris: OECD Publishing, 2021.
- [7] European Commission. *DigComp 2.2: The Digital Competence Framework for Citizens*. European Commission, 2025. <https://digital-strategy.ec.europa.eu/en/library/digcomp-22>.
- [8] Cai Y, Zhang J, Yu C, et al. Artificial intelligence literacy in the era of digital intelligence: Connotation, framework and implementation paths. *Journal of Library Science in China*, 2024(4): 71-84.
- [9] Zhong B, Liu X, Yang M. What is artificial intelligence literacy? Essence, composition and evaluation system. *Journal of East China Normal University (Educational Sciences)*, 2024, 42(1): 71-84.
- [10] Davis F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 1989, 13(3): 319-340.
- [11] Vygotsky L S. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press, 1978.
- [12] Frankena W K. *Ethics*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [13] Carolus A, Koch M J, 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*. *Computers in Human Behavior: Artificial Humans*, 2023, 1(2): 100014.
- [14] Ning Y, Zhang W, Yao D, et al. Development and validation of the Artificial Intelligence Literacy Scale for Teachers (AILST). *Education and Information Technologies*, 2025, 30(12): 17769-17803.
- [15] 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. *British Journal of Educational Technology*, 2024, 55(3): 1082-1104.
- [16] Lee S C, Baby T, Vongvit R, et al. Development and validation of Generative AI Competence Scale (GenAIComp) among university students. *Technology in Society*, 2026, 84: 103059.
- [17] Li N, Liu S, Ji J. Construction of a smart literacy evaluation system for university students from the perspective of the continuum. *Information Studies: Theory & Application*, 2025, 48(1): 150-159.
- [18] Zhu S, Li H, Wu D, et al. Constructing artificial intelligence core literacy for minors: Research on technical specifications of artificial intelligence for minors (Part II). *e-Education Research*, 2023, 44(6): 15-20.
- [19] Liu M, Ma S, Wang J. University students' learning with generative artificial intelligence: Model construction and evaluation index system. *Modern Educational Technology*, 2025, 35(11): 14-22.

- [20] Zhang J. AI-empowered foreign language learning literacy of university students: Framework construction and scale development. *Technology Enhanced Foreign Language Education*, 2025(1): 69-78.
- [21] Lee S, Park G. Development and validation of ChatGPT literacy scale. *Current Psychology*, 2024, 43(21): 18992-19004.
- [22] 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. *International Journal of Artificial Intelligence in Education*, 2025, 35(1): 398-438.
- [23] 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. *Computers and Education: Artificial Intelligence*, 2024, 7: 100282.
- [24] Yabana Kiremit B. Validity and reliability of the Turkish version of the artificial intelligence literacy scale among healthcare professionals. *International Journal of Human-Computer Interaction*, 2025, 42(10): 7087-7098.
- [25] Lintner T. A systematic review of AI literacy scales. *npj Science of Learning*, 2024, 9(1): 50.
- [26] Wang Y, Wu H, Derakhshan A, et al. Development and validation of an attitude scale toward generative AI-assisted English learning for Chinese university students. *Journal of Beijing International Studies University*, 2024, 46(6): 30-45.
- [27] Wang Y, Derakhshan A, Ghiasvand F. EFL teachers' generative artificial intelligence (GenAI) literacy: A scale development and validation study. *System*, 2025, 133: 103791.
- [28] Wang L, Fan H, Cao F, et al. Generative artificial intelligence empowering the digital literacy of international Chinese language teachers: Assessment framework, current status and suggestions. *Journal of Tianjin Normal University (Social Science)*, 2024(3): 72-83.
- [29] Walter Y. Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 2024, 21(1): 15.
- [30] Li Y, Zhu Y, Sun D, et al. Differences in the use of generative artificial intelligence in typical research scenarios: The influence of disciplinary background and artificial intelligence literacy. *Modern Distance Education Research*, 2025, 37(2): 92-101+112.
- [31] Celik I. Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 2023, 138: 107468.
- [32] UNESCO. AI and education: Guidance for policy-makers. Paris: UNESCO Publishing, 2021.