AI Literacy in Higher Education:

A Systematic Approach to Questionnaire Development and Validation

Maria Ranieri
https://orcid.org/0000-0002-8080-5436
University of Florence, Italy

Gabriele Biagini
https://orcid.org/0000-0002-6203-122X
University of Florence, Italy

Stefano Cuomo
https://orcid.org/0000-0003-3174-7337
University of Florence, Italy

ABSTRACT

This paper presents the development, refinement, and validation of the Critical Artificial Intelligence Literacy Scale, an instrument designed to measure artificial intelligence literacy across four dimensions: knowledge-related, operational, critical, and ethical. The initial version of the questionnaire, based on a robust theoretical framework and expert consultation, included 40 items and was tested with 57 doctoral students. It demonstrated strong psychometric properties (comparative fit index = 0.946, Tucker-Lewis index = 0.92) but showed limitations such as item redundancy (α = 0.947) and low performance of general items. To address these issues, the questionnaire was refined to a concise 24-item version. The revised instrument was evaluated using a sample of 314 first-year student teachers. Exploratory and confirmatory factor analyses confirmed a four-factor structure, with each dimension demonstrating strong reliability (Cronbach's alpha ranging from 0.838 to 0.912) and excellent model fit indices (comparative fit index = 0.960, root mean square error of approximation = 0.0441). The results validate the Critical Artificial Intelligence Literacy Scale as a reliable and efficient tool for assessing artificial intelligence literacy in educational settings.

KEYWORDS

Artificial Intelligence, AI Literacy, Scale Development, Assessment, Higher Education

INTRODUCTION

Artificial intelligence (AI) is rapidly permeating multiple domains of daily life (Chiu et al., 2024; Kong et al., 2024). According to the most recent edition of the United Nations' "Activities on Artificial Intelligence Report" (2023), approximately 400 projects across the United Nation system are based on the use of AI applications, with initiatives ranging from forecasting food crises and measuring water efficiency to locating schools via satellite data and implementing programs on HIV/AIDS. As AI technologies become increasingly present in our societies, the need for individuals to understand their features and potential—not only in terms of technical aspects but also through the critical evaluation of their broader societal implications—is becoming ever more relevant. In this regard, AI literacy,

generally understood as the ability to comprehend, critically assess, and responsibly engage with AI technologies, is seen by both international bodies (European Commission, 2018; Organisation for Economic Co-operation and Development, 2021; United Nations Educational, Scientific and Cultural Organization, 2021) and researchers (Biagini et al., 2024; Gašević et al., 2023; Selwyn, 2022) as a key element in responding to new challenges through the development of individuals' awareness of AI and by encouraging a more informed approach to the digital transformation of our lives. In particular, Miao et al. (2021) highlighted AI literacy as a key component of digital competence frameworks, emphasizing the necessity for equipping individuals with the ability to interact with AI ethically and effectively. Similarly, the Organisation for Economic Co-operation and Development (2021) underscored the importance of AI literacy in fostering informed decision-making, critical thinking, and responsible digital citizenship. The European Commission (European Commission, 2018) further supported this perspective by integrating AI literacy within its digital education policies, advocating that AI-related skills are essential for both students and professionals in the 21st century. Beyond institutional perspectives, academic research has extensively discussed the societal and educational implications of AI literacy. For instance, Long and Magerko (2020) highlighted that AI literacy is a crucial competency for the future workforce, as it enables individuals to collaborate effectively with AI systems rather than being displaced by them. Gašević et al. (2023) further emphasized the need for AI literacy to be an integral part of education, arguing, that without adequate understanding, individuals may struggle to navigate AI-driven environments responsibly. These contributions collectively reinforce the notion that fostering AI literacy is fundamental for preparing individuals to engage meaningfully and critically with AI technologies, thereby supporting a more informed and conscious approach to digital transformation. Focusing on the educational sector, we are witnessing the proliferation of guidelines on AI for academic teaching and learning (Jiao et al., 2024). Generally, these guidelines aim to provide guidance on the appropriate uses of AI, mostly generative AI, for educational purposes, and AI literacy is commonly mentioned as a fundamental strategy to sustain aware uses of AI applications (Biagini, 2025; Sabzalieva & Valentini, 2023). Essentially, in educational contexts, AI literacy is crucial to prepare teachers and students to critically adopt AI technologies for teaching and learning. From the teachers' perspective, AI literacy enables educators to incorporate AI tools in pedagogically meaningful ways while maintaining control over their courses (Tlili et al., 2023; United Nations Educational, Scientific and Cultural Organization, 2024). For instance, a teacher may allow students to use generative AI for brainstorming and developing preliminary ideas for a project but prohibit its use for drafting the final submission. Additionally, educators may require students to provide short descriptions of how they use AI in their work to ensure transparency and alignment with course policies. Teachers can also leverage AI for instructional design, such as in creating lecture slides or infographics, provided they verify the accuracy and fairness of AI-generated content before presenting it to students (Ding et al., 2024; Sperling et al., 2024). From the students' perspective, AI literacy is equally essential in ensuring responsible and ethical AI use in their learning processes. For instance, if an instructor allows students to use AI applications only for generating initial ideas, students should not employ these applications to fully produce their assignments. Moreover, students bear full responsibility for verifying the accuracy of AI-generated content, ensuring that it aligns with academic integrity principles and does not constitute plagiarism (Dwivedi et al., 2023). To illustrate, a student using AI to generate a preliminary draft of a report must critically assess the AI output and refine it based on academic sources rather than submitting the text as-is. Misuse of AI for academic dishonesty, such as using it to answer quizzes or bypass critical thinking tasks, constitutes e-cheating and is subject to institutional disciplinary policies (Sullivan et al., 2023). AI literacy must also encompass inclusivity and accessibility, ensuring that all students, including those with disabilities or financial constraints, can equitably engage with AI tools. For example, an instructor using AI-generated materials must verify their compatibility with screen readers to accommodate visually impaired students. Similarly, AI-generated videos should include subtitles to support students with hearing impairments. On the student side, those with special educational needs may request

permission to use AI-powered assistive tools, such as concept map generators, as part of their learning accommodations (Villarreal et al., 2023). Furthermore, when AI-based learning activities require paid software, students should have access to equivalent free alternatives to ensure equal participation in coursework. For teachers and students to be able to use AI in an aware, critical, and ethical manner, AI literacy represents an enabling factor that deserves attention from educational institutions.

However, to systematically address the promotion of AI literacy in education, tools intentionally designed to diagnose the educational needs of teachers and students are necessary. Measuring users' knowledge and attitudes towards AI across its various dimensions is a preliminary condition for designing and delivering effective AI literacy curricula for different target groups. As we will show in the next section, while there are some AI assessment tools, there is a lack of tools specifically dedicated to assessing AI literacy in its multidimensional aspects, including theoretical and practical knowledge, critical attitudes, and ethical implications. This paper aims to address this need through the presentation of an original assessment tool called the Critical AI Literacy Scale (CAILS), which, like many other tools measuring digital-related skills, literacy, and competences, is based on self-reported knowledge and attitudes (see, for example, the well-known Hargittai's [2005] study on the assessment of internet skills or the largely used tool for evaluating teachers' digital competences, namely SELFIEforTEACHERS from the European Commission, [2017]). In particular, this paper illustrates the development and validation process of this tool across three main phases. The first phase concentrated on the definition of the AI literacy framework and a preliminary analysis of existing tools for generating an initial pool of items. The second phase involved evaluating the instrument's performance and testing its items, while the third phase focused on refining the tool through further item selection and rigorous validation. Before detailing the methodology that guided the development process and presenting the results obtained, we provide an overview of the study's background.

BACKGROUND

Several tools have been developed to assess various aspects of AI literacy, reflecting the increasing relevance of AI in education, society, and professional contexts. These tools address diverse target groups, focusing on middle and secondary school students, adults, and the general population. They assess different dimensions of AI literacy, such as cognitive understanding, ethical reasoning, practical skills, and affective attitudes, through performance-based and self-reporting instruments (Biagini, 2024). Below, we grouped these tools by target audience to highlight their characteristics and inform the design of a comprehensive AI literacy assessment tool.

The AI Literacy Concept Inventory (AI-CI) (Zhang et al., 2022) is a performance-based scale designed for middle school students. It evaluates AI-related knowledge through 20 multiple-choice items derived from an AI literacy curriculum. The validation process employed the item response theory, providing strong evidence for content, structural validity, and internal consistency. Its general item content makes it adaptable across different AI contexts and less sensitive to rapid technological developments.

Another instrument targeting secondary school students is the AI Literacy Questionnaire (AILQ) (Ng et al., 2023). This self-report tool adopts a multidimensional approach, combining cognitive, affective, behavioral, and ethical dimensions into 32 five-point Likert items. Confirmatory factor analysis (CFA) identified four learning domains, offering moderate content validity and strong structural validity, though evidence for responsiveness remains limited.

The AI Literacy Scale by Kim and Lee (2022) is tailored to middle school students. It explores six key factors: societal impact, understanding AI, execution plans, problem-solving, data literacy, and ethics. Despite its robust structural validation and internal consistency, its availability is limited to Korean speakers, restricting its global use.

The Meta AI Literacy Scale (MAILS; Carolus et al., 2023) is a comprehensive tool targeting adults. It employs 34 eleven-point Likert items to evaluate AI literacy across modular domains, including

application, ethics, self-efficacy, and competency. While its modularity allows partial use, content validation is limited, and some items exhibit floor and ceiling effects.

The Scale for Non-Experts' AI Literacy (SNAIL; Laupichler et al., 2023) also focuses on adult populations, including non-experts, university students, and medical students. This self-report tool uses a three-factor model: technical understanding, critical appraisal, and practical application. Its validation processes include exploratory factor analysis and revalidation in multiple languages, yet it lacks direct cross-cultural evidence and exhibits floor effects in its responses.

The Perceived Artificial Intelligence Literacy Questionnaire (PAILQ-6; Grassini, 2024) is a concise six-item scale designed for adults. It uses a seven-point Likert format to evaluate awareness and engagement literacy. The validation process shows strong internal consistency and highlights demographic predictors, such as education and gender, on perceived AI competency.

For the general population, the AI Literacy Scale (Wang et al., 2022) focuses on human-AI interaction through 12 seven-point Likert items divided into four factors: awareness, use, evaluation, and ethics. While its structural and construct validity are high, cross-cultural validation is missing, and evidence for reliability is moderate.

Another self-report tool for the general population is the AI Self-Efficacy Scale (AISES; Wang & Chuang, 2023), which assesses AI self-efficacy. Grounded in technology-related self-efficacy research, the 22-item scale focuses on four domains: assistance, anthropomorphic interaction, comfort with AI, and technological skills. Although CFA confirmed its factor structure, content validation remains absent, limiting its applicability.

The Scale of AI Literacy for All (SAIL4ALL; Soto-Sanfiel et al., 2024) combines true/false and Likert-scale formats across four subscales: "What is AI?," "What can AI do?," "How does AI work?," and "How should AI be used?" Despite good structural validity in some subscales, others (e.g., "What can AI do?") show poor performance, and content validation is missing.

The AI Anxiety Scale (Wang & Wang, 2022) addresses AI-induced anxiety in both expert and non-expert populations. This validated multi-item scale captures the construct of AI anxiety with evidence for reliability, content validity, and nomological validity. It bridges an important gap in understanding psychological responses to AI.

The Attitude Towards Artificial Intelligence Scale (ATAI; Sindermann et al., 2021) measures attitudes toward AI across three languages—German, Chinese, and English. Its five-item scale evaluates two dimensions: acceptance and fear of AI. The scale shows good factorial consistency across cultures and highlights the relationship between attitude and willingness to interact with AI products.

Pinski and Belian's Instrument (2023) evaluates AI literacy through 13 seven-point Likert items focusing on five factors: AI technology knowledge, human actors in AI knowledge, AI steps knowledge, AI usage experience, and AI design experience. While offering a broad perspective on AI-related competences, its validation is limited by a small sample size, reducing its robustness.

Lastly, the General Attitudes Towards Artificial Intelligence Scale (GAAIS; Schepman & Rodway, 2020) captures positive and negative attitudes toward AI. Positive subscales emphasize societal and personal utility, while negative subscales address ethical and human judgment concerns. The scale demonstrated good psychometric indices and is a valuable tool for assessing general and application-specific attitudes. Table 1 summarizes the tools reviewed.

Table 1. Summary of the AI Literacy Questionnaires Reviewed

Tool name	Authors	Type	Items	Target Population	Factor Description
AI Literacy Concept Inventory (AI-CI)	Zhang et al., 2022	Performance- based	20 multiple-choice items	Middle school students	Single factor
AI Literacy Questionnaire (AILQ)	Ng et al., 2023	Self-report	32 five-point Likert items	Secondary school students	(F1) affective learning: (F1a) Intrinsic motivation, (F1b) confidence; (F2) behavioral learning: (F2a) behavioral commitment, (F2b) collaboration; (F3) cognitive learning: (F3a) know and understand, (F3b) evaluate and create; (F4) ethical learning
AI Literacy Scale by Kim and Lee	Kim & Lee, 2022	Self-report	30 five-point Likert items	Middle school students	(F1) societal impact, (F2) understanding of AI, (F3) AI execution plans, (F4) problem-solving with AI, (F5) data literacy, (F6) AI ethics
Meta AI Literacy Scale (MAILS)	Carolus et al., 2023	Self-report	34 eleven-point Likert items	Adults	(F1) AI literacy: (F1a) use & apply AI, (F1b) understand AI, (F1c) detect AI, (F1d) AI ethics; (F2) create AI; (F3) AI self-efficacy: (F3a) AI problem solving, (F3b) AI learning; (F4) AI self-competency: (F4a) AI persuasion literacy, (F4b) AI emotion regulation
Scale for Non-Experts' AI Literacy (SNAIL)	Laupichler et al., 2023	Self-report	31 seven-point Likert items	Adults	(F1) technical understanding, (F2) critical appraisal, (F3) practical application
Perceived AI Literacy Questionnaire (PAILQ-6)	Grassini, 2024	Self-report	6 seven-point Likert items	Adults	(F1) awareness literacy, (F2) engagement literacy
AI Literacy Scale (AILS)	Wang et al., 2022	Self-report	12 seven-point Likert items	General population	(F1) awareness, (F2) use, (F3) evaluation, (F4) ethics
AI Self-Efficacy Scale (AISES)	Wang & Chuang, 2023	Self-report	22 seven-point Likert items	General population	(F1) assistance, (F2) anthropomorphic interaction, (F3) comfort with AI, (F4) technological skills
Scale of AI Literacy for All (SAIL4ALL)	Soto-Sanfiel et al., 2024	Performance-based	56 true/false or five-point Likert items	General population	(T1) What is AI?, (T2) What can AI do?, (T3) How does AI work?, (T4) How should AI be used?
AI Anxiety Scale (AIAS)	Wang & Wang, 2022	Self-report	21 five-point Likert items	General population	(F1) learning, (F2) job replacement, (F3) sociotechnical blindness, (F4) AI configuration
Attitude Towards Artificial Intelligence (ATAI)	Sindermann et al., 2021	Self-report	5 five-point Likert items	General population	(F1) acceptance, (F2) fear
Pinski & Belian's Instrument	Pinski & Belian, 2023	Self-report	13 seven-point Likert items	General population	(F1) AI technology knowledge, (F2) human actors in AI knowledge, (F3) AI steps knowledge, (F4) AI usage experience, (F5) AI design experience
General Attitudes Towards AI Scale (GAAIS)	Schepman & Rodway, 2020	Self-report	20 five-point Likert items	General population	(F1) perceived capability of AI for tasks involving human judgment, (F2) perceived capability of AI for tasks involving big data

Note. AI = artificial intelligence.

Although the existing AI literacy scales are valuable and useful to start addressing the issue of measuring AI literacy, they also present some limitations. Most of them fall short of capturing the multifaceted nature of AI competences, leaving significant gaps in their scopes, applications, and designs. Indeed, they focus narrowly on specific aspects of AI literacy, such as factual knowledge,

affective attitudes, or technical skills, without integrating these dimensions into a unified framework. These limitations can be categorized into three primary areas: conceptual fragmentation, validation gaps, and contextual limitations.

A primary concern is the fragmented conceptualization of AI literacy evident in many tools. Several instruments adopt a narrow focus, concentrating on specific facets rather than embracing the multidimensional nature of AI competence, which encompasses theoretical understanding, practical application, critical evaluation, and ethical engagement. For instance, performance-based tools like the AI-CI (Zhang et al., 2022), while useful for gauging AI-related knowledge in middle school students, primarily address a single cognitive factor. Similarly, scales, such as the ATAI (Sindermann et al., 2021) and the GAAIS (Schepman & Rodway, 2020), offer insights into affective responses like acceptance or fear, but do not substantially explore users' technical understanding, critical appraisal, or practical skills for using AI. Even tools with a broader declared scope may not fully integrate these dimensions. This piecemeal assessment is problematic because an individual's ability to navigate the complexities of AI effectively depends on an interplay of these varied competencies; a narrow assessment fails to provide a holistic diagnostic picture necessary for tailoring targeted educational interventions.

Beyond the scope, significant concerns arise from weaknesses in validation processes and psychometric properties reported for several instruments. The trustworthiness of any assessment hinges on robust evidence of its validity and reliability, yet gaps are apparent. For example, the absence of thorough content validation for tools like the MAILS (Carolus et al., 2023), the AISES (Wang & Chuang, 2023), and the SAIL4ALL (Soto-Sanfiel et al., 2024) raises fundamental questions about whether these instruments accurately and comprehensively measure the intended constructs of AI literacy. Without this assurance, the interpretability and usefulness of the assessment outcomes are compromised. Furthermore, some tools exhibit specific psychometric issues: MAILS and the SNAIL (Laupichler et al., 2023) show floor or ceiling effects, potentially limiting their ability to differentiate among individuals at the extremes of the literacy spectrum. SAIL4ALL also reports poor performance in some subscales, which can undermine the validity of its dimensional scores. Even when factor structures are confirmed, as with the AILQ (Ng et al., 2023), limited evidence for responsiveness can hinder its utility in tracking changes in AI literacy levels over time or following educational programs. The robustness of validation is also impacted by sample characteristics, such as the small sample size noted for Pinski and Belian's Instrument (2023).

Finally, the generalizability and applicability of some tools are constrained. The AI Literacy Scale by Kim and Lee (2022), for instance, is only available in Korean, restricting its use in international comparative studies or diverse linguistic contexts. A more widespread issue is the lack of cross-cultural validation, explicitly noted for SNAIL (Laupichler et al., 2023) and the AI Literacy Scale (AILS) (Wang et al., 2022). This omission is critical, as cultural factors can significantly influence individuals' perceptions, understanding, and interaction with AI technologies, meaning that an instrument validated in one cultural context may not be appropriate or equivalent in another.

These identified limitations—spanning fragmented conceptualizations, incomplete validation efforts, and restricted applicability—collectively highlight a significant gap in the current landscape of AI literacy assessment.

The Critical AI Literacy Scale (CAILS), presented in this paper, directly addresses these deficiencies by: (1) adopting a comprehensive, multidimensional framework (Cuomo et al., 2022) to counter construct fragmentation systematically integrating critical and ethical dimensions often marginalized in other instruments; (2) undergoing a rigorous, multi-phase validation process to ensure psychometric soundness; and (3) being designed for adaptability across diverse educational contexts. In doing so, this tool provides a robust and comprehensive approach to evaluating AI literacy, addressing the gaps left by existing instruments and advancing the field.

METHODOLOGY

Research Questions and Objectives

This study focuses on the development, refinement, and validation of a questionnaire aimed at assessing AI literacy in higher education contexts.

To guide the process, the study addresses the following research questions:

RQ1: How can we define and operationalize the conceptual dimensions of AI literacy?
RQ2: Which evaluation instrument provides the most valid and reliable measurement of AI literacy and its dimensions?
RQ2.1: What are the main items of the instrument aligned with each dimension?
RQ2.2: Are there any items in the original set that can be excluded to enhance the instruments' efficiency?

By answering these research questions, the study aims to provide a reliable comprehensive instrument for assessing AI literacy, facilitating more efficient and effective educational interventions

RESEARCH DESIGN AND PROCEDURES

To design and develop this AI literacy scale, we adopted the systematic methodology described by DeVellis (2016), which ensures rigor in scale development. The process was structured into three main phases, each addressing specific objectives essential to the development and validation of the instrument.

Steps in the Development Process

1. Definition of the construct: Grounded in a robust theoretical framework, we refined broad themes of AI competence through an extensive literature review.
2. Item generation and content validation: We formulated specific conceptual elements aligned with the identified dimensions of AI literacy and engaged with experts to ensure robustness and relevance.
3. Initial testing and exploratory analysis: A preliminary version of the instrument was tested with a development sample to evaluate internal consistency and factor structure.
4. Refinement and validation: We optimized the instrument through confirmatory analysis, improving efficiency by reducing redundancy and enhancing clarity.
5. While the overall process included four conceptual steps, not all steps were relevant to every phase, but some of them were repeated in more than one phase. Each phase focused on a subset of these steps based on its specific goals:
   - Phase One: Conceptual model and initial set of items - Focused on step 1 to define the construct and generate items.
   - Phase Two: Scale refinement and pilot test - Addressed steps 3 and 4 to test and refine the initial instrument.
   - Phase Three: Refinement and Validation - Centered on step 4 to further refine, validate, and optimize the instrument.

Phase One: Conceptual Model and Initial Set of Items

The process began with a background analysis aimed at defining the construct of AI literacy and generating an initial set of items. Key activities included:

- Literature review and framework development: Building on existing theoretical models (Cuomo et al., 2022), we conducted a systematic review of the literature to identify the critical dimensions of AI literacy. These dimensions include knowledge-related, operational, critical, and ethical aspects. The outcome was a comprehensive framework that maps AI literacy as a multidimensional construct.
- Expert consultation and content validity: To ensure the conceptual relevance and theoretical alignment of the identified dimensions, we engaged with a panel of experts in AI, education, and psychometrics. Their feedback was used to refine the sub-constructs and generate a pool of items that adequately represented each dimension.

Phase Two: Scale Refinement and Pilot Test

Subsequently, we focused on evaluating the instrument and testing its items. This phase involved the following key activities:

- Content validity: In this phase, the process was informed by expert interviews, where feedback from professionals in AI, education, and psychometrics was used to ensure the clarity and relevance of each item.
- Exploratory factor analysis (EFA): EFA was conducted to explore the underlying factorial structure of the instrument. Factor loadings and internal consistency (Cronbach's alpha) were calculated for each dimension.
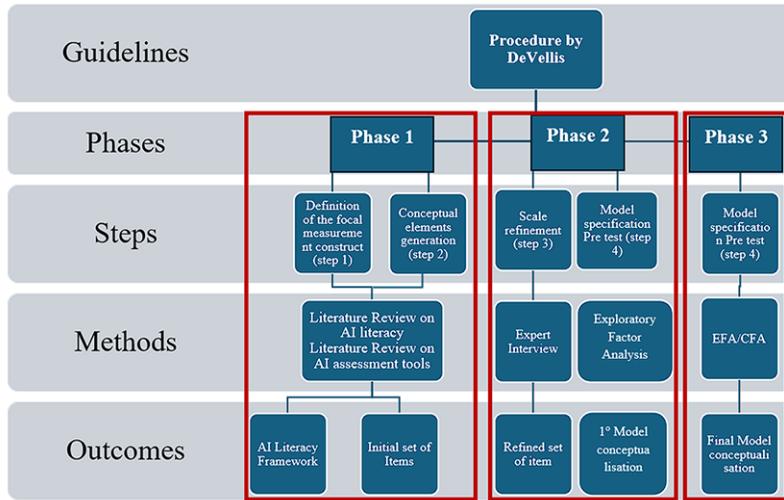
Phase Three: Refinement and Validation

The last phase aimed to refine the initial instrument through extensive testing and bigger sample, focusing on assessing its reliability and validity. Key activities included:

- EFA/CFA: Building on the outcomes of the pilot study, this phase employed both EFA and CFA to test and validate the hypothesized model. The EFA was used to reassess the factorial structure, while the CFA confirmed the model's fit and psychometric properties. These analyses culminated in a finalized version of the questionnaire, demonstrating robust reliability and validity across all dimensions.

Figure 1 summarizes the process, illustrating how the phases and steps interconnect to ensure the instrument's rigor and applicability.

Figure 1. Research Design

Note. AI = artificial intelligence; EFA/CFA = exploratory factor analysis/confirmatory factor analysis.

Ethical Considerations

This research was conducted as part of a doctoral research program at the University of Florence. According to the university's ethics guidelines, formal ethics approval was not required for this study as: (1) the data collected were completely anonymous and did not include any sensitive personal information; (2) participation was entirely voluntary; and (3) the research involved adult participants in an educational setting completing a literacy assessment questionnaire. All participants were informed about the purpose of the study and their right to withdraw at any time without consequences.

In the next paragraphs, we describe the specific procedures for each phase and related steps as well as the main outcomes of this systematic approach to questionnaire development and validation of AI literacy.

PHASE ONE: CONCEPTUAL MODEL AND INITIAL SET OF ITEMS

Definition of the Focal Measurement Construct (Step 1)

The relevant dimensions to conceptually describe the idea of AI literacy were determined after a thorough assessment of the literature. To ensure a diligent literature review, we followed established guidelines practiced by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses model (Moher et al., 2009). Aside from reliable sources like the European Commission (2018; 2019; 2020; 2021), Tuoimi (2018), and the Organisation for Economic Co-operation and Development (2018; 2025) this review also drew on seminal works by authors like Floridi (2021), Ng, and Selwyn (2022). Based on this assessment of the literature, and as discussed in the backdrop of the study, we referred to an existing framework for AI literacy (Cuomo et al., 2022) that is articulated along four dimensions: cognitive, operational, critical, and ethical. For the explicit details of the literature search procedure, including the timeframe, databases, search strings, and inclusion/exclusion criteria, please refer to Cuomo et al. (2022), where these aspects are thoroughly presented.

As anticipated, for identifying the meaningful dimensions to conceptually represent the notion of AI literacy, a thorough review of literature has been carried out, incorporating insights from seminal works such as the contributions by Floridi (2021), Ng, and Selwyn (2022), among others, and authoritative sources such as the European Commission (2018; 2019; 2020; 2021), Tuomi (2018), the

Organisation for Economic Co-operation and Development (2018; 2021; 2025), the United Nations Educational, Scientific and Cultural Organization (2021; 2019; 2021), the United Nations Children's Fund (2020; 2021);, and Dignum et al. (2021) for United Nations Children's Fund.

Through the review, we were also able to revise the different tools already adopted to evaluate the different elements of AI literacy (see "Conceptual Elements Generation [Step 2] below). On this ground, we elaborated a questionnaire to appraise participants' AI literacy by measuring their awareness across various aspects related to the cognitive, operational, critical, and ethical aspects. The framework indeed provided a structured outline for the questionnaire leading to an operationalization of the different dimensions and subdimensions, meaning that the abstract concepts were translated into specific, measurable variables or items for the questionnaire. In addition, each item was carefully crafted to be accessible to a specific public, which was higher education students.

As a first step to develop an assessment instrument about AI literacy, the definition of a conceptual model to operationalize the constructs involved was undertaken. Based on an extensive review of current research on AI literacy in a previous study (Cuomo et al., 2022), we developed a comprehensive framework that tackled the various components and interrelations at the center of AI understanding to reflect the complexity and multifaceted nature of AI literacy. The framework consisted of four essential variables that, combined, cover the entire range of AI literacy. These factors work together to create a comprehensive prism through which AI literacy can be investigated, evaluated, and developed. They stress the importance of going beyond just consuming AI passively to obtaining a more active and responsible knowledge, providing a comprehensive, integrative method to addressing AI literacy. To be more specific, the framework (see Figure 1) is made up of a:

- Knowledge-related Dimension: This dimension includes the comprehension of fundamental AI concepts, concentrating on fundamental behaviors and dispositions that do not necessitate prior technical expertise (Ng et al., 2021). It covers types of AI, machine learning fundamentals, and a range of AI applications, such as voice and vision recognition.
- Operational Dimension: This dimension emphasizes the capacity to design and implement algorithms, solve problems using AI tools, and develop simple AI applications to improve analytical and critical thinking (Kim et al., 2021). It is focused on applying AI concepts in a variety of contexts (Druga et al., 2019; Lee et al., 2021).
- Critical Dimension: This dimension emphasizes the significance of effective communication and collaboration with AI technologies as well as critical evaluation of their impact on society. It does this by highlighting AI's potential to engage students in cognitive, creative, and critical discernment activities (Su & Zhong, 2022).
- Ethical Dimension: Concerning the responsible and conscious use of AI technologies, this dimension stresses the balanced view of delicate ethical issues raised by AI, such as the delegation of personal decisions to a machine (e.g., job placement or therapeutic pathways), and emphasizes the growing attention towards "AI Ethics," encompassing transparency, fairness, responsibility, privacy, and security.

Our research advances the field of empirical research on AI literacy by building on this multidimensional approach.

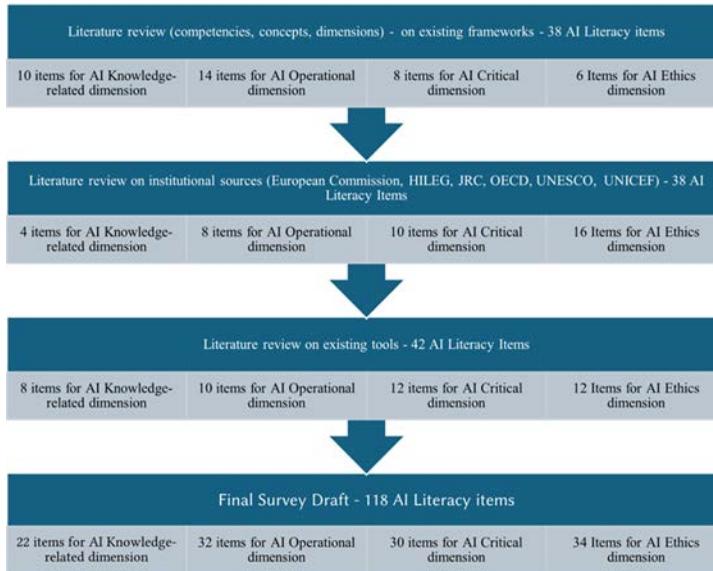## Conceptual Elements Generation (Step 2)

Based on the four dimensions identified through the framework, we developed conceptual elements for the AI literacy questionnaire using a multi-step procedure that combined literature analysis, consultation from experts, and brainstorming.

Prior to moving forward with the development of a preliminary draft of the questionnaire, we further reviewed the literature already explored to build our AI literacy framework, both research studies (N 38 items) or documents by international bodies (N = 38 items), in order to identify in the

different studies the conceptual elements to be operationalized according to the four dimensions of our AI literacy framework. To expand the item generation process, we also examined the existing questionnaires on AI literacy (see paragraph 2 above) (N = 42 items). It was only after that we were able to create a final survey draft that could address a variety of AI-related knowledge, skills, attitudes, and behaviors that are pertinent in the quickly changing technological landscape of today.

A preliminary collection of 118 AI literacy items was constructed, distributed into the four dimensions (i.e., knowledge [N = 22 items], operational [N = 32 items]; critical [N = 30 items]; ethical [N = 34 items]), and relevant phrases were gathered from multiple sources and compared, as shown in Figure 2, while Figure 3 shows the results of the initial scale.

Figure 2. Item Dimensions and Sources

Note. AI = artificial intelligence; HILEG = High-level expert group; JRC = Joint Research Center; OECD = Organisation for Economic Co-operation and Development; UNESCO = United Nations Educational, Scientific and Cultural Organization; UNICEF = United Nations Children's Fund.

Figure 3. Graphical Process for Item Generation

Framework Dimension	Description	Sample question	Matrix option	Nr. of items	References
Knowledge-related Dimension	Know how to use AI applications and its fundamental workings.	When it comes to AI, I feel my knowledge on the subject would be:	Know and understand AI definitions and theoretical foundations	22	Cuomo et al, 2022; Ng et al, 2021; UNESCO, 2019a, 2019b, 2021.
			Know and understand AI basic mathematical functions behind the algorithm		
Operational Dimension	Using AI concepts, expertise, and applications in various contexts.	In your opinion the following tasks could be supported by AI?	Supporting Emergency services	32	Cuomo et al 2022, Ng et al, 2021, Wang & Wang; JRC, 2018; OECD, 2018a, 2018b; UNESCO, 2019.
			News reporting		
			Emotional support		
Critical Dimension	AI applications for critical thinking abilities (such evaluating, appraising, predicting, and designing)	How much do you agree with the following statements?	Artificially intelligent systems make many errors.	30	Schepman & Rodway, 2022; Selwyn, 2022; Wang et al 2022; OECD, 2019; UNICEF, 2020.
			An artificially intelligent agent would be better than an employee in many routine jobs.		
Ethical Dimension	Human-centered factors (such as justice, responsibility, openness, ethics, and safety).	How much do you believe the following considerations affect the trustworthiness of AI?	Social Impact: the risk that AI will further concentrate power and wealth in the hands of the few.	34	European Commission 2018,2019,2020,2021; Floridi 2018; Floridi et al, 2021; JRC & OECD, 2021; Sindermann et al, 2021; UNICEF, 2021a, 2021b; UNESCO, 2021.
			Democratic impact: the impact of AI technologies on democracies.		
			Work impact: Impact of AI on the labour market and how different demographic groups might be affected.		

Note. AI = artificial intelligence.

PHASE TWO: SCALE REFINEMENT AND PILOT TEST (STEPS 3 AND 4)

Objectives and Methods

The primary objective of this phase was to refine the initial item pool and establish a preliminary model for measuring AI literacy. This was achieved through a systematic process combining theoretical refinement and empirical testing. The phase began with an in-depth review of the initial item pool, drawing upon the theoretical framework established in the previous phase. A pilot test was conducted to assess the internal structure and reliability of the scale, with specific attention given to identifying redundancies, ambiguities, and gaps in content coverage. EFA was employed as a statistical method to evaluate the underlying dimensions and provide initial insights into the instrument's validity and reliability.

Expert Consultation and Content Validity (Step 3)

The refinement process was informed by expert interviews, where feedback from professionals in AI, education, and psychometrics was used to ensure the clarity and relevance of each item. This phase also included an EFA, conducted with a sample of first-year student teachers, to explore the factorial structure of the instrument and assess its internal consistency. This process resulted in a refined set of items aligned with the theoretical framework and a preliminary conceptualization of the model, laying the groundwork for subsequent validation. We achieved this with open-ended expert interviews, as recommended by Moore and Benbasat (1991). Moreover, to find additional dimensions and elements of general AI literacy, as well as to get input on the initial set of issues, we conducted open-ended expert interviews. These interviews were aimed at AI specialists with a

variety of backgrounds and specialties since we took a general approach to AI literacy. We sought the aid of a group of specialists (N = 5) in the fields of AI and educational assessment to ensure the face validity of the questionnaire. The expert panel comprised:

- One full-time professor with a philosophical background and expertise in AI ethics
- One associate professor of pedagogy with experience in educational technology
- One professor of engineering specializing in AI and machine learning
- One doctoral candidate with expertise in data science and psychometrics
- One senior researcher with experience in AI applications and self-regulation

These experts were selected based on their diverse backgrounds and complementary expertise, ensuring comprehensive coverage of the AI literacy construct. It is important to note that this study focused on a cognitive task that did not require an in-depth understanding of the phenomenon being studied, so the use of a small group of experts for content validity assessment was deemed appropriate (Anderson & Gerbing, 1991; Hinkin, 1998; Schriesheim et al., 1993). These professionals had a strong knowledge of the intended construct of the questionnaire and were well-versed in AI literacy. We gave them a draft of the questionnaire and asked for their input on the items' clarity, applicability, and relevance.

The definitions were provided to each expert to guarantee that they all understood the four AI literacy constructs. The following steps made up the content approval procedure. Each item was carefully examined by the expert panel, who also offered insightful comments and recommendations for improvement. Any items that were ambiguous, repeated, or unrelated to the construct being measured were called out. Their input was crucial in helping to improve the questionnaire and make sure it accurately reflected AI literacy. Following the methodology promoted by Schriesheim et al. (1993), the experts were initially asked to classify each item into one of the four of Cuomo et al.'s (2022) constructs (i.e., AI Knowledge, AI Applications, AI Critical Assessment, and AI Ethics). An item was deemed to clearly address a topic if at least four out of the five experts classified it in the same way. There were 118 total items; of those, 15 had incorrect or unclassified classifications from two experts, while 23 had classification errors from multiple experts. These components were therefore excluded from the study. Fourteen items were rephrased, 20 items related to the impact of AI in education were moved outside the main body of the questionnaire, and they became an appendix that can be used in educational contexts as a wider information section. This procedure left us with 60 items that were then improved, including their phrasing and format, as per the experts' suggestions.

Model Specification Pre-test (Step 4): The Participants

We conducted a pre-test study with 57 participants, doctoral students, and a convenience sample, therefore neither probabilistic nor representative of the reference population, who were participating in an AI literacy course. Recruitment was conducted on a voluntary basis, and no incentives were offered for participation, and students were informed that their participation or non-participation would not affect their course grades. The participants were asked to rate their level of agreement with the general construct items and dimension items on a 5-point Likert scale ranging from "None at all" to "A great deal." Participants were chosen from the XXXVIII cycle of the Italian National Doctorate through the AI Literacy course offered by the Science of Education University of Florence.

The participants' ages ranged from 18 to 65, with the following distribution: 18-24 years (3 respondents, accounting for approximately 5.3%), 25-34 years (22 respondents, representing approximately 38.6%), 35-44 years (14 respondents, constituting approximately 24.6%), 45-54 years (15 respondents, making up approximately 26.3%), and 55-65 years (3 respondents, representing approximately 5.3%). The gender distribution was 37 (64.9%) female respondents, 18 (31.6%) male respondents, and 2 (3.5%) respondents who preferred not to disclose their gender. Among the respondents, the distribution of the highest degree of education completed was as follows:

university degree (24 respondents, accounting for approximately 42.1%), master's degree (25 respondents, representing approximately 43.9%), doctorate degree (4 respondents, constituting approximately 7.0%). We deemed the sample adequate because our construct intended to assess broad AI literacy. Although the sample size was at the lower end of the suggested range, we thought it adequate for a pre-test of internal measurements.

Model Specification Pre-Test (Step 4): Results

Step 4 involved formalizing the measurement model and conducting a pre-test study in accordance with predetermined standards (DeVellis, 2016). By examining internal consistency and item loadings, we then evaluated the generic AI literacy measurement paradigm (Fornell & Larcker, 1981).

Our initial model with four dimensions showed an overall good fit ($R^2 = .72$) and the reliability and validity of our survey were evaluated using Cronbach's alpha, McDonald's omega (1999), the composite reliability (CR), and the average variance extracted (AVE). Table 2 displays the results. The Cronbach's alpha for the survey was 0.937, and the scores for the four constructs, AI knowledge, AI applications, AI critical assessment, and AI ethics, were 0.902, 0.922, 0.910, and 0.911, respectively. The overall score of 0.937 indicated that the instrument as a whole was more reliable than the individual constructions, even though the reliabilities of all four constructs were higher than the 0.70 threshold. Fornell and Larcker's (1981) CR and AVE criteria were used to assess the scale's convergent validity. While CR levels of 0.70 and higher are considered to be good, the AVE, that compares a construct's variation to the variance brought on by measurement error is considered appropriate, demonstrating appropriate convergence at values higher than 0.5 (Hair et al., 1998). According to our scale, adequate convergence was indicated by CR values greater than 0.8 and AVE values greater than 0.5. Finally, item loadings were all above the recommended threshold of 0.70 at a significance level of $p < 0.001$. Overall, the results indicated strong empirical support for the measurement model.

Table 2. Results of Cronbach's Alpha, McDonald's Omega (1999), Average Variance Extracted (AVE), and Composite Reliability (CR)

Framework Dimensions	Cronbach's α	McDonald's ω (1999)	Average Variance Extracted (AVE)	Composite Reliability (CR)	N. of Elements
AI Knowledge	0.903	0.911	0.536	0.902	8
AI Applications	0.906	0.914	0.501	0.922	12
AI Critical Assessment	0.911	0.918	0.506	0.910	10
AI Ethics	0.916	0.921	0.511	0.911	10
Total	0.947	0.952	0.512	0.937	40

PHASE THREE: REFINEMENT AND VALIDATION (STEP 4)

Objectives and Methods

The third phase aimed at further refinement and advanced validation of the instrument. Building on the outcomes of phase two, this phase employed both EFA and CFA to test and validate the hypothesized model. The EFA was used to reassess the factorial structure, while the CFA confirmed the model's fit and psychometric properties. The sample for this phase was split, with EFA conducted on one half and CFA on the other half to ensure proper validation procedures. These analyses culminated in a finalized version of the questionnaire, demonstrating robust reliability and validity across all dimensions.

Model Specification Pre-Test (Step 4): The Participants

The participants in this phase were 314 first-year student teachers enrolled in the Primary Education program at the University of Florence. This sample was selected based on convenience, as these students represent a relevant population for assessing AI literacy in an educational context. The sample size of 314 participants satisfies established guidelines for factor analysis. According to Hair et al. (2019), a sample size greater than 200 is considered adequate for CFA, whilst Worthington and Whittaker (2006) recommend a participant-to-item ratio of at least 10:1. With 24 items in our final instrument, our sample of 314 participants (ratio of 13:1) exceeds both recommendations, ensuring robust statistical power for our analyses.

The questionnaire was administered as part of a course on AI literacy offered by the university, ensuring that participants had some exposure to the subject matter. Recruitment was conducted on a voluntary basis no incentives were offered, and students were clearly informed that participation was voluntary and would not affect their course grades.

The actual data collection occurred online in October 2023 using the survey tool "Qualtrics" with all analyses performed using the statistical software R. The survey was administered to a convenience sample of first-year student teachers at the University of Florence's Primary Education program. After excluding incomplete responses, the final sample comprised 314 student teachers, including 293 females (93.31%), 16 males (5.09%), 3 individuals who preferred not to say (0.96%), and 2 non-binary/third gender individuals (0.64%).

The age distribution of the sample was as follows: 18-24 years (73.57%, 231 respondents), 25-34 years (18.15%, 57 respondents), 35-44 years (5.41%, 17 respondents), 45-54 years (2.55%, 8 respondents), and 55-65 years (0.32%, 1 respondent). Regarding employment at the school level, 23.85% (26 individuals) were employed at early childhood education schools, 76.15% (83 individuals) at primary schools, and 1.83% (2 individuals) at high schools.

In terms of educational attainment, 74.52% (234 respondents) had a high school diploma, 15.61% (49 respondents) held a three-year university degree, 7.01% (22 respondents) possessed a five-year university degree, 1.91% (6 respondents) had a 1° level master's degree, 0.64% (2 respondents) had a 2° level master's degree, and 0.32% (1 respondent) had a doctorate.

Professional experience varied among respondents: 79.82% (87 individuals) had less than five years of experience, 15.60% (17 individuals) had 5 to 10 years, 2.75% (3 individuals) had 10 to 20 years, and 1.83% (2 individuals) had more than 20 years.

Model Specification Pre-Test (Step 4): Results

In our survey, we assessed reliability using Cronbach's alpha, McDonald's omega (1999), CR, and AVE. These findings are detailed in Table 4. The overall Cronbach's alpha for the survey was 0.914, with the four constructs, AI knowledge, AI applications, AI critical assessment, and AI ethics, scoring 0.912, 0.859, 0.838, and 0.847, respectively. Although each construct's reliability exceeded 0.70, the overall instrument's reliability was higher at 0.914, indicating greater overall consistency.

Convergent validity was assessed using CR and AVE, as per Fornell and Larcker's (1981) criteria. CR values of 0.70 or higher are deemed satisfactory, and AVE values above 0.5 indicate good convergence. In our scale, CR values were all above 0.7, and AVE values exceeded 0.5, demonstrating acceptable convergence as shown in Table 3.

Table 3. Results of Cronbach's Alpha, McDonald's Omega (1999), AVE, and CR

Framework Dimensions	Cronbach's α	McDonald's ω (1999)	Average Variance Extracted (AVE)	Composite Reliability (CR)	Number. of Elements
Knowledge- related dimension	0.912	0.914	0.672	0.924	6
Operational dimension	0.859	0.863	0.501	0.855	6
Critical dimension	0.838	0.842	0.506	0.858	6
Ethical dimension	0.847	0.850	0.502	0.856	6
All the items	0.914	0.920	0.545	0.966	24

Note. AVE = average variance extracted; CR = composite reliability.

The foundational structure of the 24-item measure was reaffirmed through EFA. Factor loadings indicated the dimensions being assessed by the survey questions, with questions targeting similar indicators loading onto the same factors. Factor loadings can range between -1.0 and 1.0. Principal component analyses were conducted, revealing a four-component structure consistent with the hypothesized framework. These components were rotated using varimax, an orthogonal rotation technique, to permit correlations among the components. Table 4 holds the details for the factor loadings.

Principal component analyses results indicated that the four variables with eigenvalues greater than 1.00 accounted for 72.28% of the total variance. This study adhered to five criteria for item retention or elimination: (1) eigenvalue >1.00; (2) factor loadings <0.50; (3) significant loadings on multiple factors; (4) at least three items per factor; and (5) exclusion of single-item factors. Ultimately, 24 items were retained from the original 40, with six items each focusing on AI knowledge, operational, critical, and ethical dimensions.

The results of the EFA are presented in Table 5. The CFA of the model with the 24 items loaded onto the four factors showed acceptable fit indices: comparative fit index (CFI) = 0.960, Tucker-Lewis index (TLI) = 0.954, root mean square error of approximation (RMSEA) = 0.0441, standardized root mean square residual (SRMR) = 0.0549. According to the results, the $\chi2/df$ value was 1.36, showing the suitability of the research model, as detailed in Table 6, while the path analysis of structural equation model can be seen in Figure 3.

Table 4. Results of Exploratory Factor Analysis

	Factor Loadings			
	Knowledge-related dimension (KW)	Operational dimension (OP)	Critical dimension (CR)	Ethical dimension (ET)
KW1	0.825			
KW2	0.872			
KW3	0.912			
KW4	0.808			
KW5	0.845			
KW6	0.624			

Table 4. Continued

	Factor Loadings			
	Knowledge-related dimension (KW)	Operational dimension (OP)	Critical dimension (CR)	Ethical dimension (ET)
OP1		0.685		
OP2		0.767		
OP3		0.612		
OP4		0.631		
OP5		0.906		
OP6		0.598		
CR1			0.655	
CR2			0.705	
CR3			0.883	
CR4			0.664	
CR5			0.652	
CR6			0.681	
ET1				0.687
ET2				0.855
ET3				0.609
ET4				0.804
ET5				0.651
ET6				0.609

Table 5. Model Fit Statistics

				RMSEA 90% CI	
CFI	TLI	SRMR	RMSEA	Inferior	Superior
0.960	0.954	0.0549	0.0441	0.0211	0.0573

Note. The four-factor model is the theoretical model. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; TLI = Tucker-Lewis index. A TLI and CFI greater than 0.900 and RMSEA values less than 0.050 suggest adequate model fit.

Figure 4. Path Analysis of Structural Equation Model

* refer to likert scale described in the methodology

DISCUSSION

The aim of this study was to develop and validate a measure to assess AI literacy in higher education to respond to the increasing demand to equip students with skills and information on AI (Gašević et al., 2023; Selwyn, 2022). Our approach was set in three additional phases that examined, refined, and validated the multifaceted construct of AI literacy according to the four most influential dimensions expressed within the literature (Cuomo et al., 2022): knowledge, functional, critical, and moral. The strategy described herein was set with the foundation from Calvani et al.'s (2009) vision for digital literacy, since their notion was the foundation upon which Cuomo et al. (2022) developed the AI Literacy model. We conducted a scoping review as per DeVellis' (2016) guidelines, picking suitable items (n = 118) that were AI literacy relevant. The item pool was reduced to 60 items based on expert comments and further condensed to 24 final items by EFA and CFA.

Our study produced a number of findings. First, the multidimensional nature of AI literacy was confirmed using EFA and CFA. The four-dimensional model—knowledge-related, operational, critical, and ethical—performed extremely well with our hypothesized model, demonstrating that these dimensions comprehensively represent the construct of AI literacy. The high reliability coefficients for each dimension (0.912, 0.859, 0.838, and 0.847, respectively), as well as for the overall instrument (0.914), confirm the internal consistency of the questionnaire. These findings align with established frameworks in the field. The knowledge-related and operational dimensions correspond closely to the AI4K12 framework's five big ideas about AI, particularly their emphasis on perception, representation, and learning as foundational AI concepts (Touretzky et al., 2019). The AI4K12 initiative identifies core concepts that K-12 students should understand about AI, including how computers perceive the world, how they learn from data, and how they interact with humans. Our knowledge-related dimension captures similar foundational understandings adapted for higher education contexts, while our operational dimension aligns with AI4K12's focus on problem-solving and the practical application of AI concepts. Similarly, our critical and ethical dimensions reflect the International Society for Technology in Education standards for educators' focus on digital citizenship and responsible technology use (International Society for Technology in Education, 2017), extending these concepts specifically to AI contexts. This convergence with internationally recognized frameworks strengthens the theoretical grounding of CAILS while offering a more comprehensive integration of ethical and critical dimensions often underrepresented in existing instruments.

Second, the validation process highlighted the importance of addressing several aspects of AI literacy. The operational aspect, which deals with practical skills in using and developing AI tools, and the critical aspect, which involves examining the social implications of AI, were particularly important. These findings are in alignment with the previous research supposing the necessity for a mixed strategy that combines technical competencies and critical thinking abilities (Kim et al., 2021; Ng et al., 2021). Moreover, during the process of refinement of the scale for the questionnaire, it was observed that some items that were drawn from other tested questionnaires, such as the GAAIS had extremely bad performance on recent implementations. This was due to the fact that, over time, questions that were based on initial opinions have not kept up with appropriate conditions and users' perceptions. We therefore eliminated such items and made the total to ensure the questionnaire would be more reliable and valid.

The practical implications of CAILS are best illustrated through implementation scenarios. Consider a typical pre-service teacher education program implementing CAILS at the beginning of an AI literacy course. The assessment might reveal that students have moderate theoretical knowledge about AI but significant gaps in ethical understanding and practical application skills. Based on this diagnostic profile, educators could design targeted interventions addressing specific dimensional weaknesses rather than delivering generic AI literacy content. For instance, if CAILS assessment reveals low scores on the ethical dimension, the intervention could focus on case studies of algorithmic bias in educational settings, privacy considerations in student data usage, and frameworks for responsible AI implementation in schools. Conversely, if the operational dimension shows weaknesses, the program could emphasize hands-on workshops with AI tools relevant to educational practice. This dimension-specific approach enables efficient, targeted skill development based on actual needs rather than assumptions about what students might lack.

Looking at existing questionnaires employed to assess AI literacy, the novelty and strength of our questionnaire lie in its integrated approach to the multi-dimensionality of AI literacy. Existing scales (Schepman & Rodway, 2022; Sindermann et al., 2021; Wang & Wang, 2022) are primarily centered on specific or individual dimensions of AI (e.g., affective or collaborative ones) or were designed for assessing AI literacy following a course. On the other hand, our poll strictly acknowledges and quantifies AI literacy's latent richness with a welcoming multilateral framework. The new focus not only fills in a critical literature gap but opens up new doorways for policymakers, educators, and researchers, alike, to build increased and holistic AI literacy across people and disciplines. With the

application of this multidimensional tool, stakeholders will be able to better understand the wide scope of AI literacy, which will enable improved educational interventions and policymaking for fostering responsible and knowledge-based use of AI technologies.

LIMITATIONS AND FUTURE DIRECTIONS

Despite the strengths of our study, several limitations should be acknowledged.

First, while the participants represented the entire population of first-year student teachers enrolled in the Primary Education program at the University of Florence, this single-institution sample limits the generalizability of findings to broader populations. The gender distribution (93% female) accurately reflects the enrolled student population but may not represent other educational contexts or disciplines.

Second, the reliance on self-reported data introduces potential response bias. Participants' perceptions of their AI literacy may not fully reflect their actual competencies. Future research could benefit from integrating objective measures of AI skills alongside self-reported assessments to provide a more comprehensive understanding of AI literacy.

Third, while our study provides initial validation evidence, cross-cultural validation remains limited. Although the instrument incorporates items from internationally validated questionnaires spanning from North America to Korea, and while we are currently including CAILS in several European research projects to address this limitation, further systematic cross-cultural validation studies are needed to ensure the instrument's applicability across different cultural and linguistic contexts.

Fourth, the current validation was conducted with students in education programs. Validation with other populations, including STEM students, in-service teachers, and professionals from other fields, would strengthen the instrument's applicability.

Finally, while the CAILS focuses on stable conceptual structures that should remain relevant over time, the rapidly evolving nature of AI technology means that periodic reviews will be necessary. We plan to conduct regular reviews of the instrument to ensure its continued relevance, though we anticipate that the fundamental dimensions will remain stable given their conceptual rather than technical focus.

Acknowledging these limitations, we propose a clear roadmap for future research:

- Enhance external validity: Future studies should continue to validate the CAILS across diverse populations, testing for measurement invariance across different academic disciplines (e.g., STEM vs. humanities) and educational levels.
- Establish criterion-related validity: To address the limitation of self-report data, the CAILS should be correlated with objective, performance-based assessments of AI skills.
- Conduct longitudinal studies: Longitudinal research is needed to track the development of AI literacy over time and to assess the effectiveness of educational interventions.
- Instrument maintenance and updates: AI is a rapidly evolving field. We plan to periodically review the CAILS to ensure its items remain relevant.

However, it is important to note that the scale was designed to assess foundational conceptual structures of AI literacy (e.g., understanding bias, critical evaluation) rather than specific, fast-changing technical knowledge. This focus on core principles is intended to give the instrument greater long-term stability and resilience against rapid technological obsolescence.

CONCLUSION

The CAILS developed and validated in this study offers a robust and versatile tool for assessing AI literacy in higher education, addressing the critical need for comprehensive evaluation instruments in this emerging field. By incorporating cognitive, operational, critical, and ethical dimensions, the questionnaire provided an in-depth measure of students' AI literacy, capturing not only technical skills but also the broader competencies needed to critically engage with AI in society. The refinement process, which reduced the questionnaire from 40 to 24 items, improved both its clarity and usability, making it more practical for educational contexts while maintaining its comprehensive scope.

The psychometric analyses conducted demonstrate that the refined instrument possesses strong construct validity and internal consistency. This validates its application as an effective tool for identifying areas where students may need additional support in developing AI-related skills and ethical awareness. In practical terms, educators can use this instrument to guide curriculum development, evaluate the effectiveness of AI literacy programs, and foster a more responsible and informed approach to AI among students.

However, as AI technologies continue to evolve at a rapid pace, future research should focus on expanding and adapting the questionnaire to ensure its ongoing relevance. This includes testing the instrument in different educational and cultural contexts, as well as exploring its applicability to other age groups and professional fields. Furthermore, longitudinal studies could provide valuable insights into how AI literacy develops over time and how educational interventions can best support this development.

In conclusion, this study provides a validated, reliable, and efficient tool that supports educators and researchers in promoting AI literacy, a critical skill for navigating an AI-driven world. Continued refinement and expansion of the questionnaire will ensure that it remains a valuable resource for fostering a responsible and inclusive integration of AI in society.

FUNDING STATEMENT

COMPETING INTERESTS STATEMENT

The authors of this publication declare there are no competing interests.

PROCESS DATES

CORRESPONDING AUTHOR

Correspondence should be addressed to Gabriele Biagini [gabriele.biagini@unifi.it], Maria Ranieri [pmaria.ranieri@unifi.it], Stefano Cuomo [stefano.cuomo@unifi.it]

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *The Journal of Applied Psychology*, *76*(5), 732–740. DOI: 10.1037/0021-9010.76.5.732

Biagini, G. (2024). Assessing the assessments: Toward a multidimensional approach to AI literacy. *Medical Education*, *15*(1), 91–101. DOI: 10.36253/me-15831 PMID: 38279173

Biagini, G. (2025). Towards an AI-Literate future: A systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education*. Advance online publication. DOI: 10.1007/s40593-025-00466-w

Biagini, G., Cuomo, S., & Ranieri, M. (2024). Less knowledge, more trust? Exploring potentially uncritical attitudes towards AI in higher education. *Italian Journal of Educational Technology*, *32*(1), 37–52. DOI: 10.17471/2499-4324/1337

Calvani, A., Fini, A., & Ranieri, M. (2009). Valutare la competenza digitale. Modelli teorici e strumenti applicativi. *Italian Journal of Educational Technology, 17*(3), 39-39.DOI: 10.1016/j.chbah.2023.100014

Carolus, , AKoch, , M. JStraka, , SLatoschik, , M. EWienrich, , 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. DOI: 10.1016/j.chbah.2023.100014

Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, *6*(6), 100171. DOI: 10.1016/j.caeo.2024.100171

Cuomo, S., Biagini, G., & Ranieri, M. (2022). *Artificial intelligence literacy, che cos'è e come promuoverla. Dall'analisi della letteratura ad una proposta di framework*. Media Education., DOI: 10.36253/me-13374

Devellis, R. F. (2016). *Scale developmen: Theory and applications* (4th ed.). SAGE Publications.

Dignum, V., Penagos, M., Pigmans, K., & Vosloo, S. (2021, Nov 19). *Policy guidance on AI for children*. Www.unicef.org

Ding, A.-C. E., Shi, L., Yang, H., & Choi, I. (2024). Enhancing teacher AI literacy and integration through different types of cases in teacher professional development. *Computers and Education Open*, *6*, 100178. DOI: 10.1016/j.caeo.2024.100178

Druga, S., Vu, S. T., Likhith, E., & Qiu, T. (2019). Inclusive AI literacy for kids around the world. *Proceedings of FabLearn*, *2019*, 104–111. DOI: 10.1145/3311890.3311904

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*(102642), 102642. Advance online publication. DOI: 10.1016/j.ijinfomgt.2023.102642

European Commission. (2018). *High-level expert group on AI*. Digital-Strategy.ec.europa.eu

European Commission. (2019, Apr 5). *Pilot the assessment list of the ethics guidelines for trustworthy AI - FUTURIUM - European Commission*. FUTURIUM - European Commission; European Union. https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/register-piloting-process-0.html

European Commission. (2020, Feb 19). *White paper on artificial intelligence: A European approach to excellence and trust*. Commission.europa.eu

European Commission. (2021). *European Strategy for Data*. Digital-Strategy.ec.europa.eu

European Union. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Union. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

European Union. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

Floridi, L. (2021). *Introduction – The importance of an ethics-first approach to the development of AI*. 1–4. DOI: 10.1007/978-3-030-81907-1_1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. DOI: 10.1007/s11023-018-9482-5 PMID: 30930541

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research*, *18*(1), 39–50. DOI: 10.1177/002224378101800104

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *4*, 100130. DOI: 10.1016/j.caeai.2023.100130

Grassini, S. (2024). A psychometric validation of the PAILQ-6: Perceived artificial intelligence literacy questionnaire. *Nordic Conference on Human-Computer Interaction, 35*, 1–10. DOI: 10.1145/3679318.3685359

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning Emea. https://eli.johogo.com/Class/CCU/SEM/_Multivariate%20Data%20Analysis_Hair.pdf

Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, *23*(3), 371–379. DOI: 10.1177/0894439305275911

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121. DOI: 10.1177/109442819800100106

International Society for Technology in Education. (2017). *ISTE standards for educators*. https://cdn.iste.org/www-root/Downloads/Downloads/Download-4070.pdf

Jiao, J., Afroogh, S., Chen, K., Atkinson, D., & Dhurandhar, A. (2024). The global landscape of academic guidelines for generative AI and large language models. *ArXiv (Cornell University)*. DOI: 10.48550/arxiv.2406.18842

Kim, S.-W., & Lee, Y. (2022). *The Artificial Intelligence Literacy Scale for Middle School Students. Journal of The Korea Society of Computer and Information, 27*(3), 225–238. DOI: 10.9708/jksci.2022.27.03.225

Kong, S.-C., Korte, S.-M., Burton, S., Keskitalo, P., Turunen, T., Smith, D., Wang, L., Chi-Kin Lee, J., & Beaton, M. C. (2024). Artificial intelligence (AI) literacy – An argument for AI literacy in education. *Innovations in Education and Teaching International, 62*(2), 1–7. DOI: 10.9708/jksci.2022.27.03.225

Laupichler, M. C., Aster, A., & Raupach, T. (2023). Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers and Education: Artificial Intelligence*, *4*, 100126. DOI: 10.1016/j.caeai.2023.100126

Lee, I., Ali, S., Zhang, H., DiPaola, D., & Breazeal, C. (2021). Developing middle school students' AI literacy. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 191–197. DOI: 10.1145/3408877.3432513

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. DOI: 10.1145/3313831.3376727

McDonald, R. P. (1999). *Test theory*. Psychology Press., DOI: 10.4324/9781410601087

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. UNESCO. DOI: 10.54675/PCSP7350

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. DOI: 10.1371/journal.pmed.1000097 PMID: 19621072

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192–222. DOI: 10.1287/isre.2.3.192

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Shen, M. Q. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*(100041). Advance online publication. DOI: 10.1016/j.caeai.2021.100041

Ng, K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2022). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, *28*(6), 8445–8501. DOI: 10.1007/s10639-022-11491-w

Organisation for Economic Co-operation and Development. (2018). Bridging the digital gender divide: Include, upskill, innovate. In *eulacfoundation.org* (pp. 1–151). OECD. https://eulacfoundation.org/system/files/digital_library/2023-07/bridging-the-digital-gender-divide.pdf

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the council on artificial intelligence*. OECD. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Organisation for Economic Co-operation and Development. (2025). *Future of Education and Skills 2030*. OECD. https://www.oecd.org/en/about/projects/future-of-education-and-skills-2030.html

Pinski, M., & Benlian, A. (2023). AI literacy - Towards measuring human competency in artificial intelligence. *Proceedings of the ... Annual Hawaii International Conference on System Sciences*, 165–175. DOI: 10.24251/HICSS.2023.021

Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and artificial intelligence in higher education: quick start guide*. IIEP Unesco - Etico | Platform on Ethics and Corruption in Education. https://etico.iiep.unesco.org/en/chatgpt-and-artificial-intelligence-higher-education-quick-start-guide

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports*, *1*(100014). Advance online publication. DOI: 10.1016/j.chbr.2020.100014 PMID: 34235291

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management*, *19*(2), 385–417. DOI: 10.1177/014920639301900208

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, *57*(4), 620–631. Advance online publication. DOI: 10.1111/ejed.12532

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2020). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI - Künstliche Intelligenz, 35*(1), 109–118. DOI: 10.1007/s13218-020-00689-0

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, *6*, 100169. DOI: 10.1016/j.caeo.2024.100169

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1), 1–10. DOI: 10.37074/jalt.2023.6.1.17

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(15), 15. Advance online publication. DOI: 10.1186/s40561-023-00237-x

Tuomi, I. (2018). *The impact of artificial intelligence on learning, teaching, and education policies for the future. JRC Science for Policy Report*. European Commission. DOI: 10.2760/12297

United Nations. (2023). *United Nations activities on artificial intelligence (AI)*. Itu.int; International Telecommunication Union (ITU). https://www.itu.int/pub/S-GEN-UNACT-2023

United Nations Children's Fund. (2020). *Policy guidance on AI for children draft 1.0*. https://www.unicef.org/innocenti/media/1326/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf

United Nations Children's Fund. (2021). *AI policy guidance: How the world responded*. https://www.unicef.org/innocenti/stories/ai-policy-guidance-how-world-responded

United Nations Educational, Scientific and Cultural Organization. (2021). *Beijing consensus on artificial intelligence and education*. https://unesdoc.unesco.org/ark:/48223/pf0000368303

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2019b). *Stepping up AI for social good*. United Nations Educational, Scientific and Cultural Organization.

Van Roy, V., Rossetti, F., Perset, K., & Galindo-Romero, L. (2021). *AI watch, national strategies on artificial intelligence: a European perspective*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/619fd0b5-d3ca-11eb-895a-01aa75ed71a1/language-en

Villarreal, R. M., Perdomo, E. V., Navarro, D. E. S., Aguilera, R. T., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, *13*(9), 1–18. DOI: 10.3390/educsci13090856

Wang, B., Rau, P.-L. P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, *42*(9), 1–14. DOI: 10.1080/0144929X.2022.2072768

Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, *29*(4), 4785–4808. DOI: 10.1007/s10639-023-12015-w

Wang, Y. Y., & Wang, Y. S. (2022). Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interactive Learning Environments*, *30*(4), 619–634. DOI: 10.1080/10494820.2019.1674887

Wenhua, Z., Hasan, M. K., Jailani, N. B., Islam, S., Safie, N., Albarakati, H. M., Aljohani, A., & Khan, M. A. (2024). A lightweight security model for ensuring patient privacy and confidentiality in telehealth applications. *Computers in Human Behavior*, *153*(108134), 108134. DOI: 10.1016/j.chb.2024.108134

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. DOI: 10.1177/0011000006288127

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education*, *33*(2), 290–324. DOI: 10.1007/s40593-022-00293-3 PMID: 35573722

Stefano Cuomo is a Research Fellow at the University of Florence specializing in educational technology and inclusive pedagogy. His research primarily focuses on Artificial Intelligence in education, particularly developing AI Literacy (AIL) as a key competence for educators and students. He explores the cognitive, operational, critical, and ethical dimensions of AI, creating conceptual frameworks for pedagogical integration. Currently investigating generative AI's potential for dialogic and personalized learning environments, his work also encompasses immersive technologies (XR) and developing open educational resources.