

# Construction of AI Literacy Evaluation System for College Students and an Empirical Study at Wuhan University

Dan Wu<sup>a,b</sup>, Xinjue Sun<sup>a</sup>, Shaobo Liang<sup>a</sup>, Chao Qiu<sup>b</sup>, Ziyi Wei<sup>c</sup>

<sup>a</sup> School of Information Management, Wuhan University, Wuhan 430072, China

<sup>b</sup> Undergraduate College of Wuhan University, Wuhan University, Wuhan 430072, China

<sup>c</sup> Faculty of Education, The University of Hong Kong, Hong Kong 999077, China

© Higher Education Press 2025

**Abstract** As artificial intelligence (AI) technology continues to evolve in the digital era, developing AI literacy among college students has become a crucial educational priority. This study aims to establish a scientific AI literacy evaluation system and to empirically assess the AI literacy levels of undergraduate students at Wuhan University, with the findings providing data support and theoretical reference for future AI education policy-making and curriculum design in higher education institutions. In response to the demands of AI education and university talent cultivation objectives, this study develops an AI literacy evaluation system for college students, based on the KSAVE (knowledge, skill, attitude, value, and ethics) model and the UNESCO AI competency framework. The system includes 4 level-1 indicators (AI attitude, AI knowledge, AI capability, and AI ethics), 10 level-2 indicators, and 25 level-3 indicators. The Delphi method was used to determine indicator content, while the analytic hierarchy process was employed to calculate the weights for each level of indicators. Through large-scale questionnaire surveys and statistical analysis, the study empirically measured the AI literacy levels of 1,651 undergraduate students at Wuhan University and analyzed variations in AI literacy across factors including gender, academic year, academic discipline, and technical background. The results demonstrate that the constructed AI literacy evaluation system is scientifically sound and highly applicable, providing a comprehensive and objective measure of students' AI literacy levels. Furthermore, notable differences were observed in AI literacy levels across different dimensions among Wuhan University undergraduates, with variables such as academic discipline, technical background, and participation in digital

intelligence education programs significantly influencing students' AI literacy, particularly in knowledge and capability dimensions.

**Keywords** artificial intelligence (AI) literacy, digital intelligence education, AI, evaluation system

## 1 Introduction

The explosive growth and development of artificial intelligence (AI) tools in recent years have revolutionized and propelled global economic, social, and technological progress. Particularly, the recent rise of large language models (LLMs) has profoundly updated human production tools, enhanced productivity, and reshaped lifestyles. The interaction between humanity and AI has sparked deep reflection on child–AI interaction (Druga et al., 2017), human–AI interaction in social settings (Pereira et al., 2014), and human–AI interaction in medical environments (Gratch et al., 2014). As a potentially disruptive technology, AI has achieved significant breakthroughs across multiple fields. The surge of generative AI tools and products, such as ChatGPT, has lowered the barriers for AI application, making AI education a critical component in the evolution of higher education.

In the current era of the revolutionary transformation of production tools, higher education is also undergoing profound changes. AI technology is being increasingly used in education, driving adjustments in traditional teaching models, subject content, and talent training objectives. The goals of higher education are no longer limited to knowledge transfer and storage; instead, there is a growing emphasis on fostering students' innovation capabilities, practical skills, and

Received January 9, 2025; revised February 5, 2025; accepted February 10, 2025

Dan Wu (✉)

E-mail: woodan@whu.edu.cn

their overall literacy to adapt to future technological changes. Additionally, more attention is being paid to the interactions between AI and multiple stakeholders, including students, teachers, and government leaders (Chiu, 2023). In this context, AI literacy has been given new significance as a crucial capability in higher education. In response to the widespread impact of AI technology, the education sector has begun to explore models for the deep integration of AI with education. The Ministry of Education of the People’s Republic of China (MoE) has issued policies calling for the integration of AI literacy into the higher education talent development system (MoE, 2024a & 2024b). In September 2024, UNESCO released the *AI Competency Framework for Students* (UNESCO, 2025a), which underscored the cultivation of AI literacy as one of the crucial education missions globally. These documents provide directional guidance for AI literacy education and highlight practical demands for developing and assessing AI literacy among college students.

Despite the strong emphasis on AI literacy cultivation at the policy level, in practice, higher education institutions still face gaps in evaluating and improving students’ AI literacy. Only a few countries are developing or implementing relevant AI education standards and curricula for students (UNESCO, 2025b). Particularly in the context of the digital intelligence era, universities urgently need an evaluation system that can reflect the multidimensional characteristics of students’ AI literacy to guide educational practice and facilitate teaching improvements. Consequently, the creation of a scientifically robust AI literacy evaluation system is of paramount importance. Such a system can help universities assess the current level of students’ AI literacy and provide valuable insights for optimizing curricula and enhancing teaching quality.

This paper aims to construct an AI Literacy Evaluation System for College Students (AILES-CS), comprehensively assessing college students’ AI literacy from multiple dimensions including knowledge, capability, attitude, and ethics. At the same time, this study not only provides universities with a quantifiable and replicable evaluation tool, but also investigates current AI literacy levels among students through empirical research. By shifting teaching strategies from experience-based approaches to evidence-driven optimization, it offers data-backed insights to elevate AI education quality. Combining tool development with empirical analysis, the research delivers science-informed guidance for cultivating AI literacy in higher education during the digital intelligence era. This dual approach enhances educational effectiveness while laying foundations for developing interdisciplinary talents required by modern society.

## 2 Literature Review

The concept of AI was first proposed in 1956 by scientists such as John McCarthy, marking the formal establishment of the discipline (Andresen, 2002). When scholars such as Agre (1972) provided an AI literature guide for technical professionals, the term “AI literacy” was first coined. Against the backdrop of expanding AI application, the limitations of expert systems began to surface, including difficulties in knowledge acquisition, constraints in reasoning methods, and a lack of learning capabilities, leading AI development into a period of stagnation in the 1980s (Partridge, 1987). Academia’s focus on AI gradually shifted from a technology-centered approach to a human-centered approach (Scammell, 2000). This transition marked a crucial first step in the use of AI technology for societal purposes, as scholars increasingly recognized that effective human–computer interaction should not only focus on the transmission of knowledge but also on the sharing of knowledge and the enhancement of capabilities and communication skills (Gill, 1991). Moreover, fueled by the advancement of information technologies such as Big Data, cloud computing, and the Internet of Things (IoT), along with improvements in computing power and algorithms, as well as the explosive development of AI tools and products, the concept of AI Literacy has become a focal point of public attention. Since 2018, the number of papers published in the field of AI literacy research has significantly increased (Tenório et al., 2023). As we approach the AI era, both academia and the education sector agree on the urgent need to enhance people’s ability to use AI (Su, 2018).

Regarding the definition of AI literacy, Kandlhofer et al. (2016) and others identified seven core themes that make up AI literacy, including automata theory, intelligent agents, graphs and data structures, sorting algorithms, search problem-solving methods, classical planning, and machine learning techniques. Wang et al. (2022) proposed a definition of AI literacy that described an individual’s ability to use AI technology, which means AI literacy is defined as the ability to correctly identify, use, and evaluate AI-related products within ethical standards. Furthermore, the content and methodology of AI literacy education have been in the spotlight. Wang et al. (2022) attempted to stimulate students’ AI thinking skills and enhance their AI literacy using the AI-assisted Bayesian network machine learning approach. Ng et al. (2021) and others highlighted the necessity of AI literacy education, arguing that AI literacy has become an essential digital competence for everyone.

Given the unique role of AI technology in today's information society, AI literacy has gradually become a core competence that various social groups should focus on. The development of AI literacy evaluation systems and measurement scales is also one of the key areas of current AI literacy research. A range of AI literacy measurement tools has emerged, such as scales and evaluation systems aimed at different groups, including the general public (Wang & Chuang, 2024), K-12 students (Casal-Otero et al., 2023), higher education students (Bewersdorff et al., 2024; Su et al., 2024), and teachers (Celik, 2023). There are also AI literacy measurement tools specific to different sectors such as healthcare (Karaca et al., 2021) and higher education. However, most of the existing AI literacy scales or indicator systems have not undergone detailed validity and reliability testing, and not all scales have been validated through practical application, moreover, many studies lack an explanation of explainable indicators (Lintner, 2024).

Based on existing scales and evaluation systems, this study also references the KSAVE Model. The KSAVE Model provides a structured assessment framework for evaluating AI literacy from five dimensions: knowledge, skills, attitudes, values, and ethics. The key competencies embedded within this model that align with 21st-century learning and professional requirements establish a solid theoretical foundation for constructing the evaluation system (Griffin et al., 2012).

In the current era of informatization and intelligence, AI literacy has gradually become a key indicator for measuring the overall quality of college students. Moreover, college students are one of the primary focuses of research on AI literacy measurement and evaluation. College students receiving systematic academic training, while also facing multiple challenges in their future careers. In this context, they often have complex thoughts and feelings about AI technology (Černý, 2024). Therefore, cultivating AI literacy holds significant importance for their academic development and social adaptation. There is an urgent need to construct an AI literacy evaluation system for college students based on their practical skills requirements in combination with existing theoretical frameworks and policy guidance.

## 3 Method

Given the above research background and literature review, this study aims to address the following research questions:

RQ1: What specific contents and dimensions should the AILES-CS include? How should the weights

of indicators at different levels be distributed? Does this evaluation system demonstrate effectiveness?

RQ2: Using the AILES-CS as a measurement tool for empirical research, what is the current AI literacy level of Wuhan University undergraduates? What characteristics and differences are exhibited across different dimensions?

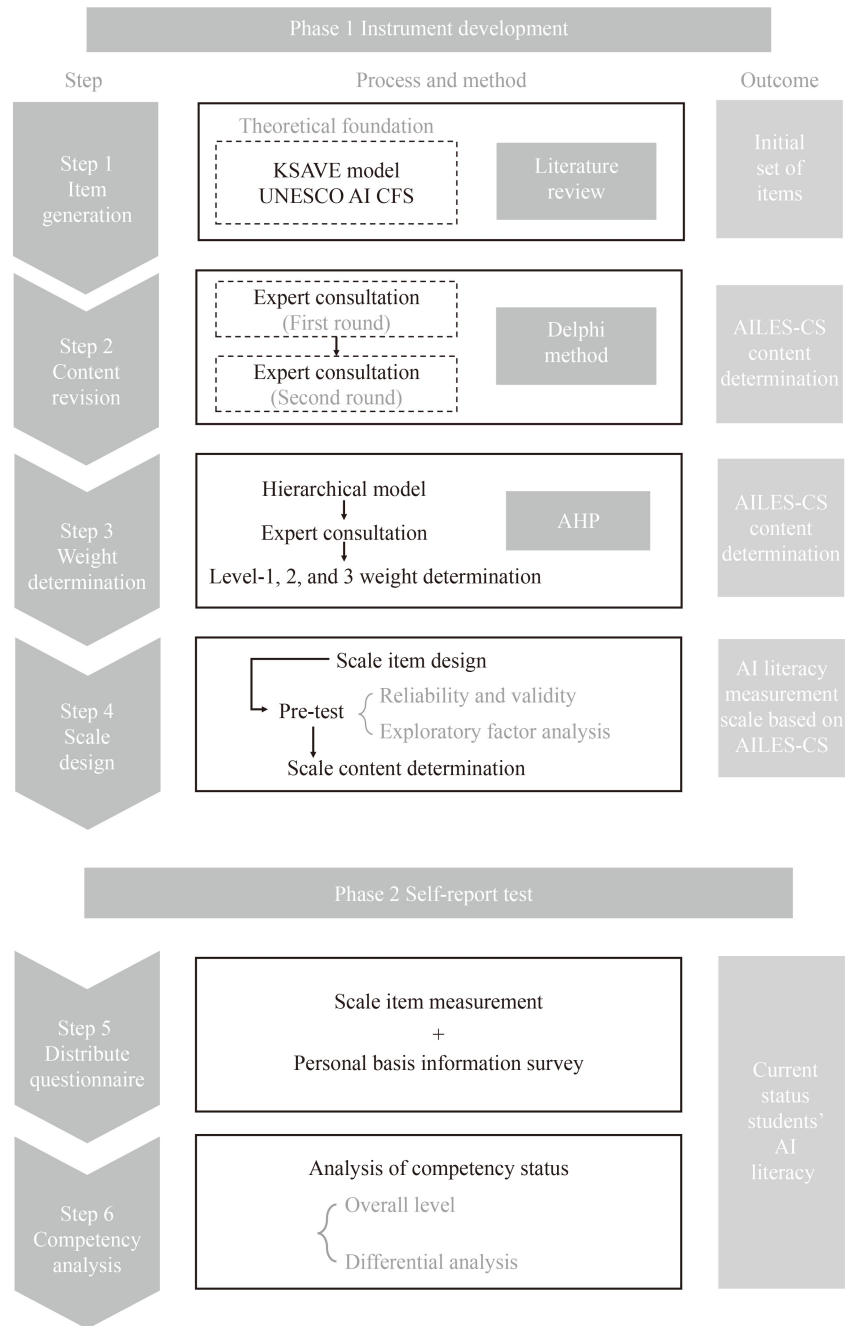
In terms of research design, this study consists of two phases: instrument development and self-report test, encompassing six steps from project integration to capability analysis. The detailed research design is shown in Figure 1.

### 3.1 | Instrument Development

The first phase, “instrument development,” aims to construct and validate a scientific AI literacy evaluation system and assessment tools. This phase is grounded in extensive literature review and internationally recognized AI literacy frameworks, including the KSAVE (knowledge, skills, attitudes, values, and ethics) model and *AI CFS*. Based on these theoretical frameworks, an initial evaluation system was constructed, which includes four level-1 indicators: “attitude,” “knowledge,” “capability,” and “ethics.” These level-1 indicators were further detailed into several level-2 and level-3 indicators, forming the complete preliminary indicator framework.

To further optimize the indicators and determine their respective weights, this study adopted the Delphi method. Experts with substantial expertise in AI-related fields were invited to revise, review, and optimize the indicator system. Through two rounds of expert consultation, experts rated and provided suggestions on the relevance, importance, and applicability of each indicator, resulting in a finalized and optimized evaluation system. Subsequently, the Analytic Hierarchy Process (AHP) was employed to determine the weights of the indicators at all levels. By constructing a hierarchical model of AI literacy evaluation indicators and developing pairwise comparison matrices, the importance of each indicator was evaluated, and the consistency of the expert consultation results was checked, ensuring the scientific and reasonable allocation of weights for the level-1, level-2, and level-3 indicators.

Based on the AILES-CS with finalized contents and weights, this study also developed an AI literacy assessment scale for college students. The scale design was guided by the level-3 indicators, using a 5-point Likert scale for scoring. The scale's items were designed to be clear and concise, supplemented with relevant examples to facilitate understanding and responding for students. To verify the scale's applicability and scientific validity, a pilot test was conducted with undergraduate and postgraduate students from different disciplines



**Figure 1** Research design.

and universities. The data from the pilot test were used to examine and optimize the scale's reliability and validity.

### 3.2 | Self-Report Test

The second phase, "self-report test," which is based on the assessment scale developed in the first phase, involves an empirical study of the AI literacy levels of undergraduate students at Wuhan University. In this phase, data were collected through a questionnaire

survey on students at the target university. Descriptive statistical analysis and differential analysis methods were employed to explore the distribution of AI literacy across different demographic groups (such as gender, discipline, and grades). With the seamless connection of these two phases, this study systematically achieves its objectives of instrument development and application validation. By scientific measurement methods, the study comprehensively reveals the current AI literacy among college students, offering both theoretical and practical support for improving AI literacy education.

## 4 Construction of AILES-CS and Development of the Scale

### 4.1 | Evaluation System Design

AILES-CS constructed in this study focuses on college students and is designed around four level-1 indicators: AI attitude, AI knowledge, AI capability, and AI ethics. The design of relevant indicators is based on the comprehensive development approach of the KSAVE model, guided by the progressive competency approach in the UNESCO (2025a), and is completed in consideration of the findings from the current literature review. Under the level-1 indicators, level-2 and level-3 indicators are designed in greater detail. For example, AI capability includes two level-2 indicators: C1—AI recognition and C2—AI application, while AI recognition encompasses two level-3 dimensions, namely C11—technology discrimination and C12—content recognition. Similarly, AI ethics includes D1—awareness of AI technology risk prevention, D2—ethical morality, and several related level-3 dimensions.

The construction of the AILES-CS has been elaborated in the team's prior research *Wuhan University AI Literacy Evaluation Guide*, part of the *Wuhan University Series on Digital Intelligence Education* (Wu et al., 2024). The evaluation system was constructed based on the KSAVE model and UNESCO (2025a) as the fundamental framework of the evaluation system. The Delphi method was employed to determine the content of the indicators, while the same method was utilized to establish the weights for each level of indicators.

The guidelines elaborate in detail on the selection criteria, weight distribution, and evaluation methods for level-1, level-2, and level-3 indicators. The overall construction of the indicator system provides theoretical support for comprehensively assessing college students' AI literacy levels, while also offering practical references for universities to optimize curriculum design and enhance the quality of AI education.

### 4.2 | Scale Development and Application Validation

Based on the AILES-CS framework, this study has developed a scale for the quantitative evaluation of college students' AI literacy. To verify the scale's applicability and reliability, this study has systematically validated the scale through item analysis, exploratory factor analysis (EFA), and reliability analysis. Item

analysis was used to select key items, ensuring that each item contributes to the overall evaluation objective. Exploratory factor analysis assessed the structural validity of the scale to verify the alignment between the items and the indicator system. Reliability analysis examined the internal consistency of the scale to ensure the stability and reliability of the evaluation results. Through these methods, a scientifically sound AI literacy evaluation tool were developed, providing quantitative support for university education practice.

To validate the scale's effectiveness, a pilot test was conducted with undergraduate and postgraduate students from science, engineering, humanities, and history programs at different universities across different provinces and cities in China. A sample of students was surveyed, with a total of 139 valid questionnaires returned. Data from the survey were analyzed using SPSS 26.0 software in terms of items, exploratory factors, and reliability.

#### 4.2.1 Item Analysis

Item discrimination measures the ability of a test item to differentiate between respondents' traits. A highly discriminative item is capable of effectively differentiating between the proficiency levels of respondents, such that individuals with greater competency tend to achieve higher scores on these items, whereas those with lesser proficiency tend to score lower.

In this study, the critical ratio method was employed to assess the reliability of the items in the scale. The critical ratio method is a data processing approach where respondents are ranked in descending order based on their total scores. In this study, respondents scoring above 134 points were classified into the high-score group, while those scoring 115 points or below were classified into the low-score group. This classification allows for more accurate evaluation of the reliability of each item in the measurement scale, providing robust support for subsequent data analysis and result interpretation.

An independent samples *t*-test was performed on the scores of the high-score and low-score groups for each evaluation item, with detailed results shown in Table 1. As seen in the table, under conditions of both equal or unequal variance, all items demonstrated significant differences, *Sig.* (two-tailed) of  $0.00 < 0.01$  and  $|t| > 3$ . This indicates significant differences between the high- and low-score groups for all evaluation items, demonstrating good item discrimination and justifying their retention in the scale.

#### 4.2.2 Exploratory Factor Analysis

When developing scale items based on the AILES-CS,

**Table 1** Independent samples *t*-test results for high- and low-score groups

Level-3 indicators	Equal variance assumption (assumed/not assumed)	Levene's test for equality of variances		<i>t</i> -test for equality of means				
		<i>F</i>	<i>Sig.</i>	<i>t</i>	Degree of freedom	<i>Sig.</i> (two-tailed)	Mean difference	Standard error difference
A11	Equal variance assumed	0.006	0.937	-8.119	87.000	0.000	-1.386	0.171
	Equal variance not assumed			-8.144	82.139	0.000	-1.386	0.170
A12	Equal variance assumed	0.804	0.372	-8.004	87.000	0.000	-1.364	0.170
	Equal variance not assumed			-8.015	86.046	0.000	-1.364	0.170
A21	Equal variance assumed	3.652	0.059	-5.423	87.000	0.000	-1.099	0.203
	Equal variance not assumed			-5.439	82.304	0.000	-1.099	0.202
A22	Equal variance assumed	10.755	0.001	-7.145	87.000	0.000	-1.127	0.158
	Equal variance not assumed			-7.188	69.125	0.000	-1.127	0.157
B11	Equal variance assumed	9.025	0.003	-6.737	87.000	0.000	-1.054	0.156
	Equal variance not assumed			-6.772	73.412	0.000	-1.054	0.156
B12	Equal variance assumed	3.879	0.052	-6.092	87.000	0.000	-1.266	0.208
	Equal variance not assumed			-6.101	86.022	0.000	-1.266	0.207
B21	Equal variance assumed	0.227	0.635	-6.306	87.000	0.000	-1.667	0.264
	Equal variance not assumed			-6.299	85.905	0.000	-1.667	0.265
B22	Equal variance assumed	1.295	0.258	-5.190	87.000	0.000	-1.298	0.250
	Equal variance not assumed			-5.183	85.451	0.000	-1.298	0.250
B31	Equal variance assumed	0.018	0.895	-6.238	87.000	0.000	-1.268	0.203
	Equal variance not assumed			-6.229	84.990	0.000	-1.268	0.204
B32_1	Equal variance assumed	14.390	0.000	-7.737	87.000	0.000	-1.392	0.180
	Equal variance not assumed			-7.786	67.942	0.000	-1.392	0.179
B32_2	Equal variance assumed	7.470	0.008	-6.825	87.000	0.000	-1.213	0.178
	Equal variance not assumed			-6.859	73.927	0.000	-1.213	0.177
B33	Equal variance assumed	5.278	0.024	-7.128	87.000	0.000	-1.257	0.176
	Equal variance not assumed			-7.163	74.596	0.000	-1.257	0.175
C11	Equal variance assumed	1.753	0.189	-6.958	87.000	0.000	-1.161	0.167
	Equal variance not assumed			-6.983	80.218	0.000	-1.161	0.166
C12	Equal variance assumed	0.394	0.532	-5.663	87.000	0.000	-1.044	0.184
	Equal variance not assumed			-5.662	86.882	0.000	-1.044	0.184
C21_1	Equal variance assumed	1.913	0.170	-8.096	87.000	0.000	-1.409	0.174
	Equal variance not assumed			-8.135	74.874	0.000	-1.409	0.173
C21_2	Equal variance assumed	42.668	0.000	-9.232	87.000	0.000	-1.640	0.178
	Equal variance not assumed			-9.305	59.667	0.000	-1.640	0.176
C22_1	Equal variance assumed	3.661	0.059	-8.031	87.000	0.000	-1.301	0.162
	Equal variance not assumed			-8.068	76.059	0.000	-1.301	0.161
C22_2	Equal variance assumed	4.655	0.034	-9.620	87.000	0.000	-1.696	0.176
	Equal variance not assumed			-9.656	79.482	0.000	-1.696	0.176
C22_3	Equal variance assumed	4.732	0.032	-7.393	87.000	0.000	-1.345	0.182
	Equal variance not assumed			-7.426	76.045	0.000	-1.345	0.181
C23_1	Equal variance assumed	13.922	0.000	-7.369	87.000	0.000	-1.280	0.174
	Equal variance not assumed			-7.422	62.958	0.000	-1.280	0.172
C23_2	Equal variance assumed	0.154	0.696	-7.131	87.000	0.000	-1.297	0.182
	Equal variance not assumed			-7.139	86.466	0.000	-1.297	0.182

(Continued)

Level-3 indicators	Equal variance assumption (assumed/not assumed)	Levene's test for equality of variances		t-test for equality of means				
		F	Sig.	t	Degree of freedom	Sig. (two-tailed)	Mean difference	Standard error difference
C31	Equal variance assumed	0.463	0.498	-8.206	87.000	0.000	-1.428	0.174
	Equal variance not assumed			-8.226	83.917	0.000	-1.428	0.174
C32	Equal variance assumed	0.078	0.780	-4.534	87.000	0.000	-1.178	0.260
	Equal variance not assumed			-4.530	86.029	0.000	-1.178	0.260
D11_1	Equal variance assumed	1.503	0.224	-5.598	87.000	0.000	-0.941	0.168
	Equal variance not assumed			-5.614	82.259	0.000	-0.941	0.168
D11_2	Equal variance assumed	7.022	0.010	-6.692	87.000	0.000	-1.037	0.155
	Equal variance not assumed			-6.729	71.839	0.000	-1.037	0.154
D12	Equal variance assumed	11.514	0.001	-7.053	87.000	0.000	-1.142	0.162
	Equal variance not assumed			-7.086	74.947	0.000	-1.142	0.161
D13	Equal variance assumed	1.205	0.275	-4.837	87.000	0.000	-0.787	0.163
	Equal variance not assumed			-4.843	86.418	0.000	-0.787	0.162
D21_1	Equal variance assumed	0.172	0.679	-5.619	87.000	0.000	-0.748	0.133
	Equal variance not assumed			-5.632	84.330	0.000	-0.748	0.133
D21_2	Equal variance assumed	1.436	0.234	-4.224	87.000	0.000	-0.613	0.145
	Equal variance not assumed			-4.232	85.299	0.000	-0.613	0.145
D22	Equal variance assumed	3.368	0.070	-4.988	87.000	0.000	-0.834	0.167
	Equal variance not assumed			-5.009	77.777	0.000	-0.834	0.167
D23	Equal variance assumed	4.870	0.030	-5.636	87.000	0.000	-0.860	0.153
	Equal variance not assumed			-5.657	80.018	0.000	-0.860	0.152
D24	Equal variance assumed	2.915	0.091	-5.273	87.000	0.000	-0.816	0.155
	Equal variance not assumed			-5.310	64.081	0.000	-0.816	0.154

Note. Sig. = Significance.

existing literature and relevant literacy evaluation scales were referenced for improvements, ensuring the scale content validity. To further validate the questionnaire's validity, the KMO test and Bartlett's test of sphericity were conducted, with the results shown in Table 2. The KMO value reached 0.839, significantly higher than the conventional threshold of 0.8, indicating strong correlations between the variables. Additionally, Sig. = 0.000, which is well below 0.05, further confirmed the existence of common factors among the data. Based on these results, the dataset is suitable for exploratory factor analysis.

Next, exploratory factor analysis was conducted on the valid sample data to extract influencing factors and explore the consistency between the AI literacy evaluation scale for college students and its corresponding evaluation indicator system structure. The analysis of AI attitude showed a KMO value of 0.621, indicating that the dataset is suitable for further exploratory factor analysis. Detailed results are shown in Table 3. To better explain the reasonableness of the scale items, principal component analysis was

employed, with two principal components extracted from the original four items. These two components explained 76.04% of the variance, demonstrating their capacity to explain the items adequately. Component 1 represents "willingness to accept AI," and Component 2 represents "emotional judgment of AI," with the detailed factor loading matrix shown in Table 4.

**Table 2** Pilot test sample test

KMO	Bartlett's test of sphericity		
	Approximate chi-square	Degree of freedom	Sig.
0.839	2717.712	528.000	0.000

Note. Sig. = Significance.

**Table 3** AI attitude test

KMO	Bartlett's test of sphericity		
	Approximate chi-square	Degree of freedom	Sig.
0.621	160.252	6.000	0.000

Note. Sig. = Significance.

**Table 4** Total variance explained for AI attitude

Component	Initial eigenvalue			Extraction sum of squared loadings			Rotation sum of squared loadings		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
1	2.217	55.436	55.436	2.217	55.436	55.436	1.623	40.585	40.585
2	0.824	20.607	76.043	0.824	20.607	76.043	1.418	35.458	76.043
3	0.641	16.035	92.078						
4	0.317	7.922	100.000						

According to [Table 5](#), Item A22 exhibits factor loadings close to 0.5 in both Component 1 and Component 2, indicating a high correlation. Specifically, the correlation for Component 2 is 0.569, slightly higher than the correlation of 0.491 for Component 1. Given that students might have a vague understanding of the concept of collaboration, the item was rephrased as “I am willing to accept AI’s help to solve problems or complete tasks, such as using generative AI in work or study,” to better reflect college students’ attitudes toward AI.

According to [Table 6](#), the KMO value for AI knowledge is 0.739, indicating its suitability for factor analysis. Therefore, exploratory factor analysis was conducted on AI knowledge. Through principal component analysis, three main components were extracted, explaining 76.04% of the variance, which exceeds the 60% threshold, demonstrating that these components adequately explain the items. Based on the constructed evaluation system, Component 1 was defined as understanding of basic AI concepts, component 2 as AI technology knowledge, and Component 3 as cognition of AI application Development ([Table 7](#)).

According to [Table 8](#), since Item B12 shows significant factor correlations with both Component 1 and Component 3, and the principles of AI implementation may be vaguely interpreted by students as needing technical mastery, it was revised to “I understand that AI is achieved through data analysis, machine learning algorithms, and other techniques” and categorized under Component 1 (basic AI concepts). Item B31, with factor loadings greater than 0.5 in both Components 1 and 3, was further analyzed and categorized under “history of AI development.” Consequently, it was revised to “I know that AI has

**Table 7** Total variance explained for AI knowledge

Component	Initial eigenvalue			Extraction sum of squared loadings			Rotation sum of squared loadings		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
1	3.193	39.912	39.912	3.193	39.912	39.912	2.287	28.591	28.591
2	1.588	19.850	59.762	1.588	19.850	59.762	2.085	26.059	54.650
3	0.955	11.934	71.696	0.955	11.934	71.696	1.364	17.046	71.696
...	...	...	...	...	...	...	...	...	...

**Table 5** Transposed component matrix for AI attitude

Item	Component	
	1	2
A11	0.694	0.306
A12	0.946	0.054
A21	0.051	0.949
A22	0.491	0.569

**Table 6** AI knowledge test

KMO	Bartlett’s test of sphericity		
	Approximate chi-square	Degree of freedom	Sig.
0.739	420.528	28.000	0.000

Note. Sig. = Significance.

developed through several stages, from theory to practice, from technical beginnings to continuous breakthroughs, and I am aware of important historical events in AI development (e.g., the Turing Test, Deep Blue defeating the world chess champion).”

For AI capability, exploratory factor analysis was conducted, revealing a KMO value of 0.875, indicating suitability for factor analysis, as shown in [Table 9](#).

Through exploratory factor analysis, the study extracted three principal components from the measurement items: Component 1—AI recognition ability, Component 2—AI application ability, and Component 3—AI innovation and creation. The detailed results are presented in [Table 10](#). These three components collectively explained 61.52% of the variance, indicating that they adequately explain the items. From the analysis, it is evident that the items are well distributed across the three factors, with factor

**Table 8** Transposed component matrix for AI knowledge

Item	Component		
	1	2	3
B11	0.909	0.058	0.195
B12	0.352	0.636	0.174
B21	-0.048	0.837	0.186
B22	0.116	0.869	-0.022
B31	0.539	0.085	0.558
B32_1	-0.027	0.363	0.701
B32_2	0.121	0.009	0.895
B33	0.265	0.029	0.826

**Table 9** AI capability test

KMO	Bartlett's test of sphericity		
	Approximate chi-square	Degree of freedom	Sig.
0.875	618.389	55.000	0.000

Note. Sig. = Significance.

**Table 10** Transposed component matrix for AI capability

Item	Component		
	1	2	3
C11	0.749	0.235	0.154
C12	0.716	0.119	0.167
C21_1	0.386	0.254	0.626
C21_2	0.362	0.680	0.106
C22_1	0.326	0.686	0.115
C22_2	0.340	0.411	0.572
C22_3	0.050	0.738	0.258
C23_1	0.126	0.810	0.038
C23_2	0.275	0.603	0.200
C31	-0.064	0.502	0.622
C32	0.112	0.040	0.857

loadings all exceeding 0.5. However, item C21\_1, initially categorized under Component 2—AI application ability, has a factor loading of 0.626 in Component 3—AI innovation and creation. As a result, this item was revised to “I can independently acquire the AI products I need.” Item C22\_2 showed strong correlations with both Component 2 and Component 3, so it was considered for removal.

For AI ethics, exploratory factor analysis was also conducted, with results showing a KMO value of 0.838. Two common factors were extracted, explaining 60.69% of cumulative variance, exceeding the 60% threshold and thus demonstrating the effectiveness of the factor extraction process (see Table 11).

The rotated factor loadings are shown in Table 12. Observing the data in the table, it is clear that all

nine items have factor loadings greater than 0.5, indicating that these items are informative and can effectively convey the intended content, hence there is no need for item removal. Factor 1's loadings are primarily concentrated on items under D1, such as D11\_1, D12, and D13, representing the “risk awareness.” Conversely, Factor 2's loadings are mainly found in items under D2, such as D21\_1 and D22, representing the “ethics.”

#### 4.2.3 Reliability Analysis

The study employed Cronbach's  $\alpha$  coefficient as a measure of the reliability of the tool. Specifically, when Cronbach's  $\alpha$  coefficient exceeds 0.8, it indicates very high reliability; when the coefficient falls between 0.7 and 0.8, the reliability is considered good. A coefficient between 0.6 and 0.7 suggests acceptable reliability, and a coefficient below 0.6 indicates insufficient reliability. As shown in Table 13, Cronbach's  $\alpha$  coefficients for all dimensions of the scale exceed 0.6, with the overall scale coefficient being 0.921, demonstrating that the

**Table 11** AI ethics test

KMO	Bartlett's test of sphericity		
	Approximate chi-square	Degree of freedom	Sig.
0.838	677.603	45	0.000

Note. Sig. = Significance.

**Table 12** Transposed component matrix for AI ethics

Item	Component	
	1	2
D11_1	0.597	0.342
D11_2	0.576	0.474
D12	0.737	0.212
D13	0.700	0.142
D21_1	0.048	0.804
D21_2	0.161	0.755
D22	0.332	0.734
D23	0.146	0.807
D24	0.359	0.651

**Table 13** Scale reliability analysis

Dimension	Cronbach's $\alpha$ coefficient	Number of items
AI attitude	0.644	4
AI knowledge	0.730	8
AI capability	0.841	11
AI ethics	0.873	9
AI literacy (overall)	0.921	32

developed scale has high reliability.

Based on the results of exploratory factor analysis, the initial version of the scale was optimized in terms of phrasing, and items that did not align with the factor structure or contributed minimally were removed, ensuring the applicability of the final scale. After the adjustments, the final scale consists of 31 items, distributed as follows: 4 items for AI attitude, 8 items for AI knowledge, 10 items for AI capability, and 9 items for AI ethics. The design of the scale items reflects the importance of each dimension, ensuring the effectiveness and reliability of subsequent measurements.

### 4.3 | Evaluation Tool

Below is the complete AILES-CS, including level-1, level-2, and level-3 indicators and their corresponding weights (Table 14), along with the AI literacy scale developed based on the AILES-CS (Table 15).

## 5 Survey on AI Literacy Abilities of Undergraduate at Wuhan University Based on AILES-CS

### 5.1 | Participants

A total of 2,201 questionnaires were collected, of which 1,651 passed the validity check, resulting in an effective response rate of 75.01%. Table 16 presents the basic demographic characteristics of the study sample. The total sample consists of 1,651 participants, covering dimensions such as gender, year of study, faculty, programming experience, and participation in digital intelligence education courses. Regarding gender, 56.21% of the participants were male and 43.79% were female, with the male proportion slightly higher. In terms of the year of study, first-year students made up the highest proportion (55.72%), followed by second-year students at 22.53%. The proportion of third-year and fourth-year students (and above) was relatively lower, at 10.54% and 11.21%, respectively. In terms of faculty, students from the Faculty of Information Science and the Faculty of Science accounted for 25.50% and 22.90%, respectively. The proportions of students from the Faculty of Humanities (19.81%), Faculty of Social Sciences (10.30%), Faculty of Engineering (10.18%), Faculty of Medicine (6.30%), and interdisciplinary fields (5.03%) were relatively lower, with detailed faculty distribution shown in Figure 2. Additionally, 68.75% of the participants reported having prior programming experience, while 31.25% had never studied programming, indicating that the participants generally have some technical

background. Furthermore, 57.84% of the participants had taken core courses in digital intelligence education, while 42.16% had not taken such courses. Overall, the sample in this study is highly representative, encompassing multi-dimensional information about Wuhan University undergraduates, such as gender, year of study, academic discipline, and technical background.

### 5.2 | Score Calculation Process

To assess the AI literacy levels of college students, a hierarchical weighted calculation method is employed based on the AILES-CS and the weights of the various indicators in the scale. This method is used in combination with the ratings for the student-targeted scale (5-point Likert scale) items to calculate students' AI literacy scores. The formula for calculating the score for each level of indicators is as follows:

$$S = \sum_{i=1}^n (X_i \cdot W_i).$$

First, the weighted score ( $S$ ) for each Level-3 indicator is calculated based on the rating data for the scale items ( $X_i$ ) and the corresponding weights ( $W_i$ ) of the level-3 indicators. Next, the level-3 indicator score under each level-2 indicator is weighted to compute the weighted score for the level-2 indicator. For example, for the level-2 indicators "A1 Willingness to accept" and "A2 Emotional judgment" under the level-1 indicator "AI attitude," the score calculation is as follows:

$$S_A = S_{A1} \cdot W_{A1} + S_{A2} \cdot W_{A2}.$$

Following this calculation approach, the final score for each student is calculated using the following formula:

$$S = S_A \cdot W_A + S_B \cdot W_B + S_C \cdot W_C + S_D \cdot W_D.$$

The final score  $S$  obtained by the above method represents the total AI literacy score for a student, with a theoretical score range from 1 to 5. A higher score indicates a higher level of AI literacy.

### 5.3 | Overall AI Literacy Level of Students

A statistical analysis was conducted on the AI literacy scores of undergraduates at Wuhan University across four dimensions: AI attitude, AI knowledge, AI capability, and AI ethics. The results revealed that students generally achieved favorable scores in each dimension. Most students' total scores fell within the range of 3.6 to 4.7, indicating a generally positive level of cognitive awareness and competence in AI-related fields (Figure 3). The high density of scores within this

**Table 14** AILES-CS

Level-1 indicators	Level-2 indicators	Level-3 indicators	Description of level-3 indicators	
A AI attitude (20.67%)	A1 Willingness to accept AI (0.458)	A11 Interest in AI technology (0.583)	Willingness to actively learn and understand AI technology.	
		A12 Tendency to explore AI technology (0.417)	Proactively discover and apply AI products in daily life.	
	A2 Emotional judgment of AI (0.542)	A21 Dialectical view on the pros and cons of AI (0.667)	Ability to objectively evaluate the positive and negative impacts of AI technology.	
		A22 Willingness to collaborate in AI usage (0.333)	Openness and initiative in collaborating with AI to solve problems or complete tasks.	
B AI knowledge (13.42%)	B1 Understanding of basic AI concepts (0.438)	B11 Definition of AI (0.542)	Understanding that AI is a technology that uses machines to simulate, extend, and enhance human perceptual intelligence.	
		B12 Principles of AI implementation (0.458)	Understanding the core operational mechanisms of AI technology, including how AI systems learn, reason, decide, and perform tasks from data.	
	B2 AI technology knowledge (0.292)	B21 Programming knowledge (0.417)	Mastery of the basic syntax of at least one programming language (such as Python, R, and Java) and related AI libraries (such as TensorFlow, PyTorch, and scikit-learn).	
		B22 Algorithm knowledge (0.583)	Familiarity with different types of machine learning algorithms (such as decision trees, support vector machines, and neural networks) and deep learning models (such as convolutional neural networks, and recurrent neural networks).	
	B3 Cognition of AI application development (0.270)	B31 History of AI development (0.208)	Understanding the development level of AI during different periods.	
		B32 Common AI applications (0.458)	Understanding the main application areas of AI and typical applications in daily life.	
		B33 AI development trends (0.312)	Understanding the main directions and trends in the future development of AI.	
	C AI capability (35.94%)	C1 AI recognition ability (0.292)	C11 AI technology discrimination (0.50)	Ability to distinguish between technologies or products that use or do not use AI, and recognize whether they are interacting with AI.
			C12 AI content recognition (0.50)	Ability to identify AI-generated content (such as AI-generated artwork, and AI-written content).
C2 AI application ability (0.416)		C21 AI acquisition ability (0.396)	Ability to acquire AI learning resources and application tools based on needs.	
		C22 AI collaboration ability (0.438)	Ability to collaborate with AI to solve problems and complete tasks.	
		C23 AI evaluation ability (0.166)	Ability to critically evaluate the results presented by AI.	
C3 AI innovation and creation (0.292)		C31 AI innovation ability (0.625)	Ability to propose new ideas, methods, and solutions based on AI.	
		C32 AI product creation (0.375)	Ability to create one's own work/product using AI technology.	
D AI ethics (29.98%)	D1 Awareness of AI technology risk prevention (0.542)	D11 Data and privacy protection (0.458)	Awareness of potential privacy violations, data leaks, and risks in AI usage, and awareness of information security and data protection.	
		D12 Technological bias (0.167)	Recognition of data bias, algorithm bias, and algorithmic discrimination that may exist in the design, development, and use of AI.	
		D13 Legal awareness (0.375)	Ability to follow relevant laws and regulations in the use of AI technology and applications.	
	D2 Ethical morality (0.458)	D21 Human-centered principle in AI usage (0.163)	Awareness that AI development must adhere to human- centered and benevolent empowerment principles.	
		D22 Moral cognition of AI (0.187)	Adherence to correct moral concepts and values during the development and use of AI.	
		D23 Sustainable development (0.375)	Awareness that AI technology must be safe, controllable, and lead to sustainable development.	
		D24 Social responsibility in AI usage (0.275)	Responsibility towards society and others, developing responsible AI.	

**Table 15** Scale based on AILES-CS

Code	Level-3 Indicators	Weight	No.	Scale Content
A11	Interest in AI technology	0.583	1	I actively explore, learn, and keep up with new AI products or technologies.
A12	Tendency to explore AI technology	0.417	2	In daily life, I actively follow cutting-edge AI news and updates.
A21	Dialectical view on the pros and cons of AI	0.667	3	I understand that the development of AI technology brings numerous benefits to human survival and development, while also acknowledging potential social controversies and risks.
A22	Willingness to collaborate in AI usage	0.333	4	I am willing to accept AI's help to solve problems or complete tasks, such as using generative AI in work or study.
B11	Definition of AI	0.542	5	I know that AI is a technology that uses machines to simulate and extend human perception.
B12	Principles of AI implementation	0.458	6	I understand that AI is achieved through data analysis, machine learning algorithms, and other techniques.
B21	Programming knowledge	0.417	7	I am familiar with the basic syntax of at least one programming language (such as Python, R, and Java) and can write simple code to implement some AI applications, such as classification or prediction tasks.
B22	Algorithm knowledge	0.583	8	I am familiar with different types of machine learning algorithms (such as decision trees, support vector machines, and neural networks) and deep learning models (such as convolutional neural networks, and recurrent neural networks), and understand their application scenarios in AI.
B31	History of AI development	0.208	9	I know that AI has developed through several stages, from theory to practice, from technical beginnings to continuous breakthroughs, and I am aware of important historical events in AI development (e.g., the Turing Test, and Deep Blue defeating world chess champion).
B32_1	Common AI applications	0.229	10	I can list various common AI products, both domestic and international, such as ChatGPT, NewBing, Wenxin Yiyan, and Midjourney.
B32_2	Common AI applications	0.229	11	I understand some common AI applications in daily life, such as facial recognition, autonomous driving, intelligent finance, intelligent healthcare, voice assistants, and many other scenarios.
B33	AI development trends	0.312	12	I know that AI is transitioning toward artificial general intelligence, and in the future, it will be widely applied in various scenarios, such as the development trends of generative AI and explainable AI.
C11	AI technology discrimination	0.500	13	I can distinguish between AI technology or products and traditional information technologies, and recognize whether I am interacting with AI.
C12	AI content recognition	0.500	14	I can identify AI-generated content, such as distinguishing whether an image is generated by AI tools or whether text is written by AI.
C21_1	AI acquisition ability	0.198	15	I can independently acquire the AI products I need.
C21_2	AI acquisition ability	0.198	16	I have used generative AI products, such as ChatGPT, Wenxin Yiyan, Doubao, and Kimi.
C22_1	AI collaboration ability	0.219	17	In study and work, I use AI tools (such as smart translation, text polishing) to improve efficiency.
C22_2	AI collaboration ability	0.219	18	I am willing to share and discuss experiences of using AI products with others.
C23_1	AI evaluation ability	0.083	19	I can maintain a critical and dialectical attitude towards preset results provided by AI.
C23_2	AI evaluation ability	0.083	20	I can maintain a critical attitude towards the reasoning process and conclusions of AI, and correct information based on feedback.
C31	AI innovation ability	0.625	21	I can use AI technology to propose innovative ideas and solutions, such as suggesting AI-based solutions for specific problems in life or work.
C32	AI product creation	0.375	22	I can creatively use open-source platforms to build simple AI applications or develop simple AI products using AI technology.
D11_1	Data and privacy protection	0.229	23	When using AI tools, I focus on privacy protection, such as not exposing personal information during interactions with AI.
D11_2	Data and privacy protection	0.229	24	I am aware that AI may bring risks such as data leakage and data bias, for example, facial recognition technology may cause information leakage risks.
D12	Technological bias	0.167	25	I understand that there may be technical biases, algorithmic biases, and algorithmic discrimination in the design, development, and use of AI, such as potential gender or racial bias in algorithms.

(Continued)

Code	Level-3 Indicators	Weight	No.	Scale Content
D13	Legal awareness	0.375	26	I comply with relevant laws and regulations when using AI products.
D21_1	Human-centered principle in AI usage	0.081	27	I understand that AI is inherently meant to serve humanity, not replace humans.
D21_2	Human-centered principle in AI usage	0.082	28	I believe that in human-AI interaction and collaboration (e.g., using AI products like ChatGPT), humans should play a leading role.
D22	Moral cognition of AI	0.187	29	I follow social ethical concepts such as fairness, justice, and integrity when using AI, and I can judge the moral boundaries in AI use, such as avoiding infringement on others' privacy or interests.
D23	Sustainable development	0.375	30	I believe that developers and users of AI should jointly ensure the safety and controllability of the technology and promote its sustainable development.
D24	Social responsibility in AI usage	0.275	31	I believe that when using or developing AI technology, the potential impact on society and others should be considered, and corresponding responsibilities should be taken.

**Table 16** Participants' demographic profile ( $N = 1,651$ )

Participants' demographic information	Options	Numbers	Proportion in the sample (%)
Gender	Male	928	56.21
	Female	723	43.79
Grade	First grade	920	55.72
	Second grade	372	22.53
	Third grade	174	10.54
	Fourth grade and above	185	11.21
Faculty	Faculty of Information Sciences	421	25.50
	Faculty of Science	378	22.90
	Faculty of Humanities	327	19.81
	Faculty of Social Sciences	170	10.30
	Faculty of Engineering	168	10.18
	Faculty of Medicine	104	6.30
Programming learning status	Interdisciplinary	83	5.03
	Have learned	1,135	68.75
Participation in the core courses of Digital Intelligence Education at Wuhan University	Have not learned	516	31.25
	Participated	955	57.84
	Not participated	696	42.16

range further suggests a balanced distribution of AI literacy among the students, with most scores clustering around the average and few extreme outliers.

Among the four dimensions, notable differences were observed. AI ethics scored the highest (mean value  $(M) = 4.362$ , standard deviation  $(SD) = 0.556$ ), reflecting students' heightened awareness and consensus on ethical issues in AI, likely influenced by increased public discourse on topics such as privacy breaches and algorithmic bias. AI attitude ( $M = 4.159$ ) and AI capability ( $M = 3.985$ ) also scored relatively

high, highlighting undergraduates' positive attitudes toward AI and their proficiency in using AI technologies and tools, likely driven by the widespread availability and accessibility of AI technologies and tools. However, AI knowledge scored the lowest (mean value  $(M) = 3.766$ , standard deviation  $(SD) = 0.668$ ), with significant individual differences. This suggests a notable gap in foundational AI knowledge, potentially due to the specialized nature of AI concepts and limited emphasis on AI education in the curriculum (see [Table 17](#) and [Figure 4](#)).

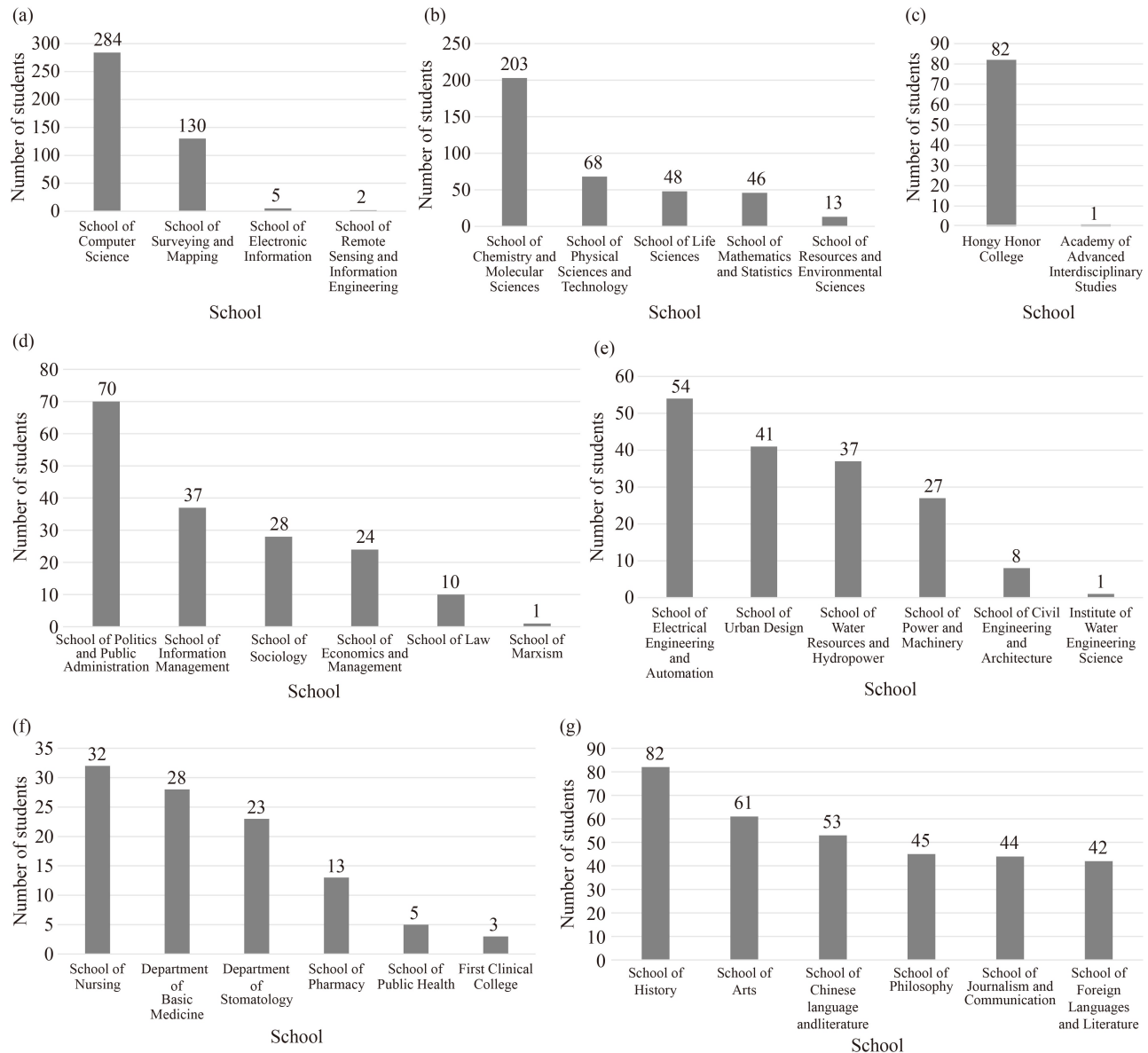
In summary, while students performed well in the dimensions of AI attitude, AI capability, and AI ethics, the lack of AI knowledge may hinder further improvements in overall AI literacy. Future initiatives should focus on enhancing AI knowledge dissemination, particularly for students from non-technical backgrounds, while maintaining attention to ethical considerations to promote the holistic and responsible development of AI.

### 5.4 | Analysis of Differences in Multi-Dimensional AI Literacy Abilities

The subsequent analysis systematically explores variations in students' AI literacy across key factors such as gender, academic year, faculty, technical background, and involvement in digital intelligence education.

#### 5.4.1 Gender Differences

An independent-samples  $t$ -test examined the influence of gender on AI literacy dimensions (AI attitude, AI knowledge, AI capability, AI ethics) and total scores in AI literacy. Results showed statistically significant differences ( $p < 0.05$ ) in AI attitude ( $p = 0.042$ ) and AI knowledge ( $p = 0.044$ ), with males scoring slightly higher than females in both dimensions. However, no significant differences ( $p > 0.05$ ) were observed for AI capability ( $p = 0.635$ ), AI ethics ( $p = 0.259$ ), or total



**Figure 2** Student statistics by faculty. (a) Faculty of Information Sciences; (b) Faculty of Science; (c) Faculty of Interdisciplinary; (d) Faculty of Social Sciences; (e) Faculty of Engineering; (f) Faculty of Medicine; (g) Faculty of Humanities.

scores ( $p = 0.47$ ).

Overall, gender had a limited impact on overall AI literacy. Males exhibited advantages in AI attitude and AI knowledge ( $p = 0.042$  vs.  $p = 0.044$ ). However, no significant differences emerged in AI capability, AI ethics, or total scores ( $p > 0.05$ ), indicating comparable performance across genders in these areas. Overall, the influence of gender on AI literacy primarily manifests in AI attitude and AI knowledge, while the impact of gender on other dimensions appears to be less pronounced (see Table 18 and Figure 5).

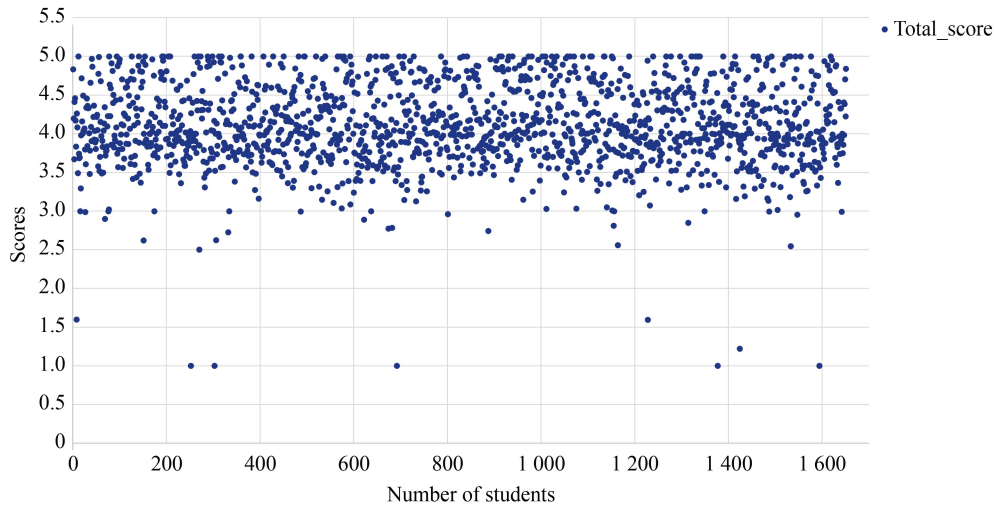
#### 5.4.2 Grade Differences

An analysis of variance (ANOVA), as shown in Figure

6, was conducted to examine differences in AI literacy of students in different academic years at Wuhan University, covering first-grade through fourth-grade students. Significant differences were found across all four dimensions and the total score of AI literacy.

First-grade students exhibited the highest overall AI literacy (total\_score =  $4.16 \pm 0.49$ ), particularly excelling in AI attitude and AI ethics. Second-grade students scored the lowest across all dimensions and total scores (total\_score =  $3.99 \pm 0.52$ ). Third-grade students (total\_score =  $4.12 \pm 0.63$ ) and fourth-grade students (total\_score =  $4.07 \pm 0.60$ ) demonstrated similar AI literacy levels.

Noteworthy is that AI literacy does not show a linear upward trajectory across academic years. First-



**Figure 3** Total score distribution of undergraduates.

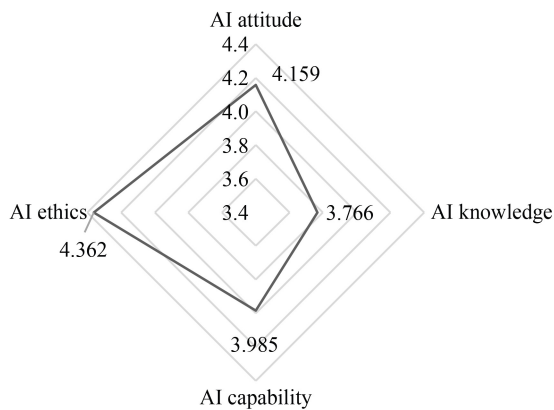
**Table 17** Total score statistics results

Indicator	Min	Max	M	SD	Median
AI attitude	1	5	4.159	0.684	4.084
AI knowledge	1	5	3.766	0.668	3.730
AI capability	1	5	3.985	0.631	4.000
AI ethics	1	5	4.362	0.556	4.392
Total_score	1	5	4.105	0.531	4.052

**Table 18** T-test results for gender differences in AI literacy-related dimensions and total scores

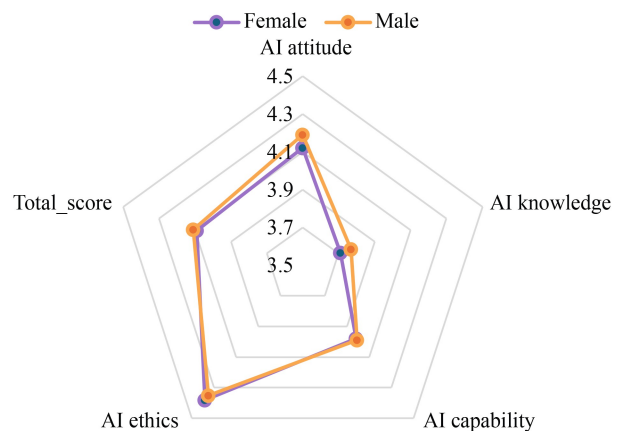
Indicator	Gender	M	SD	p
AI attitude	Female	4.12	0.66	0.042*
	Male	4.19	0.70	
AI knowledge	Female	3.71	0.64	0.044*
	Male	3.77	0.68	
AI capability	Female	3.98	0.61	0.635
	Male	3.99	0.64	
AI ethics	Female	4.38	0.55	0.259
	Male	4.35	0.56	
Total_score	Female	4.09	0.51	0.470
	Male	4.11	0.54	

Note. \* $p < 0.05$ .



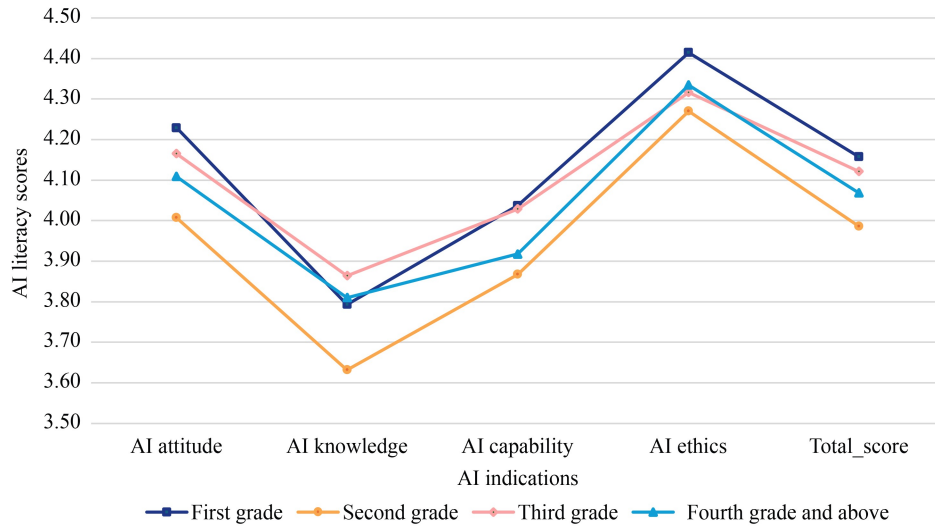
**Figure 4** Overall undergraduate AI literacy first-level indicator dimension scores.

grade students achieve significantly higher total scores and perform better in specific dimensions than their senior peers. This might be explained by factors such as sample size, course participation rates, and unique characteristics of each academic year. This could also be attributed to the strong interest in AI-related topics and higher engagement with digital intelligence courses among first-grade students. For example, more than 60% of first-grade undergraduates have taken digital intelligence courses offered by Wuhan University (as shown in Table 19). Such high engagement levels might



**Figure 5** Gender comparison of average AI literacy scores.

have a significant impact on their AI literacy levels. Overall, while AI literacy levels are generally high, there



**Figure 6** Comparison of AI literacy scores across different grades.

is no positive correlation between academic years and AI literacy scores.

**Table 19** Proportion of participants in digital intelligence education courses across different grades

Grade	Number of participants	Total number	Proportion
First grade	557	920	60.54%
Second grade	186	372	50.00%
Third grade	108	174	62.07%
Fourth grade and above	106	185	57.30%

#### 5.4.3 Faculty Differences

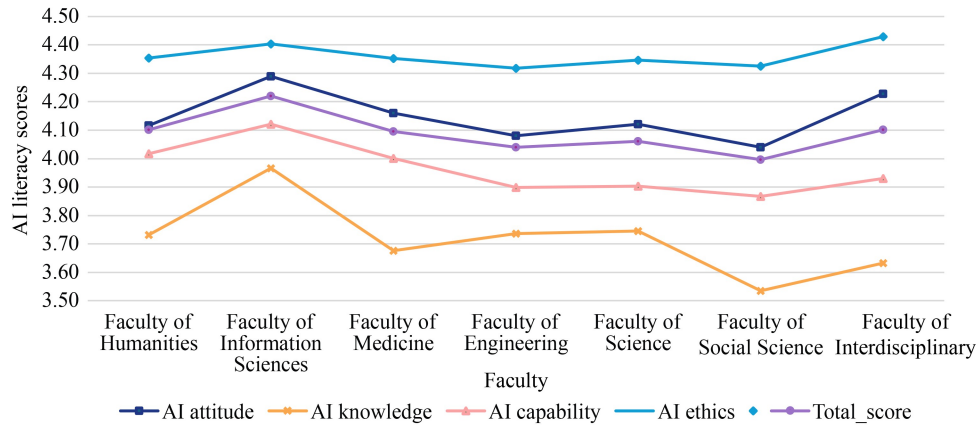
**Figure 7** summarizes the differences in AI literacy scores of students across faculties, highlighting statistical variations in different dimensions (AI attitude, AI knowledge, AI capability, AI ethics and total scores). Statistically significant differences are found in AI literacy scores in AI attitude, AI knowledge, AI capability, and total scores ( $p < 0.01$ ). However, no significant differences were observed in the AI ethics ( $p = 0.443$ ). Students from the Faculty of Information Science outperformed others, especially in AI knowledge ( $3.97 \pm 0.64$ ) and AI capability ( $4.12 \pm 0.58$ ), with the highest total score ( $4.22 \pm 0.47$ ). Students from the Faculty of Interdisciplinary Studies excelled in AI ethics ( $4.43 \pm 0.53$ ) and demonstrated strengths in AI attitude ( $4.23 \pm 0.64$ ) and AI knowledge ( $3.63 \pm 0.58$ ), the effect size measure Cohen's  $f$  for the AI knowledge dimension was 0.202, surpassing the threshold for a small effect size, reflecting their multidisciplinary advantages in AI ethics and multi-dimensional thinking. Conversely, students from the Faculty of Humanities and the Faculty of Social Sciences scored

lower in AI literacy, particularly in technical dimensions like AI knowledge ( $3.73 \pm 0.72$  and  $3.54 \pm 0.60$ , respectively) and AI capability ( $4.02 \pm 0.70$  and  $3.87 \pm 0.60$ , respectively). Their total score was also lower, with students from the Faculty of Social Sciences scoring the lowest ( $4.00 \pm 0.49$ ).

Students from the Faculty of Engineering (total\_score =  $4.04 \pm 0.57$ ), the Faculty of Science (total\_score =  $4.06 \pm 0.52$ ) and the Faculty of Medicine (total\_score =  $4.10 \pm 0.45$ ) exhibited a balanced, yet not exceptional, level of AI literacy. This may be attributed to the inclusion of AI knowledge and practices within their curricula, although not at the same technical level as in the Faculty of Information Science. Overall, students from the Faculty of Information Science and the Faculty of Interdisciplinary Studies displayed particularly strong performance in AI literacy, whereas students from the Faculty of Humanities and the Faculty of Social Sciences showed weaknesses in technical dimensions. This underscores the need for enhanced interventions in AI literacy development among undergraduates. Moving forward, it is essential to implement tailored educational strategies that account for disciplinary backgrounds, such as strengthening the technical and practical skills of students in the humanities and social sciences, further capitalizing on the advantages of interdisciplinary integration, and refining AI curriculum design to holistically elevate students' AI literacy (see **Figure 7**).

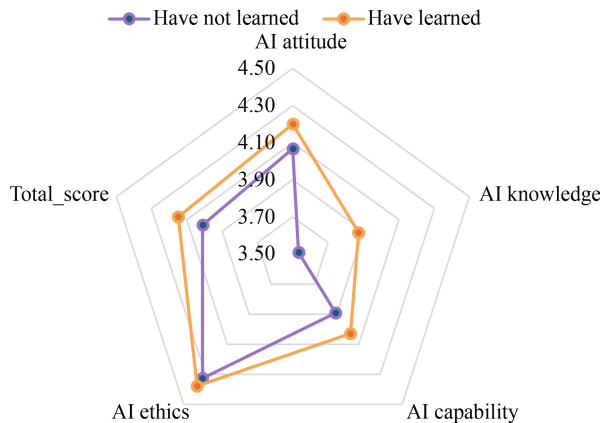
#### 5.4.4 Technical Background

As shown in **Figure 8**, independent-samples  $t$ -tests illustrate the variation in the total score and the scores in different dimensions of AI literacy among students with varying programming experience. The results reveal that students with programming experience



**Figure 7** Comparison of AI literacy scores across different faculty.

scored significantly higher than their peers without programming experience in AI attitude ( $4.20 \pm 0.67$  vs.  $4.07 \pm 0.71$ ), AI knowledge ( $3.87 \pm 0.66$  vs.  $3.53 \pm 0.62$ ), AI capability ( $4.03 \pm 0.63$  vs.  $3.89 \pm 0.63$ ), and total\_score ( $4.15 \pm 0.53$  vs.  $4.01 \pm 0.53$ ), suggesting that programming experience significantly enhances students' AI literacy (see [Figure 8](#)).



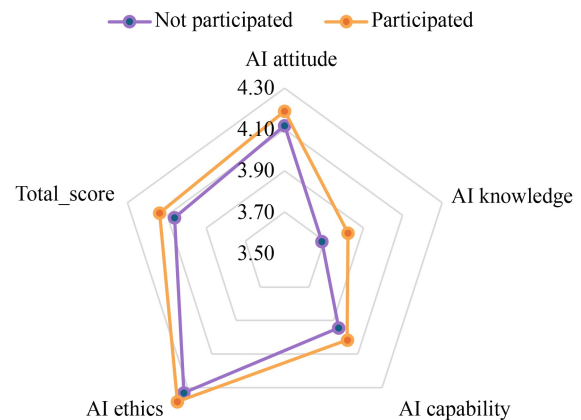
**Figure 8** AI literacy scores based on different technical backgrounds.

The most pronounced difference was found in AI knowledge ( $3.87 \pm 0.66$  vs.  $3.53 \pm 0.62$ ), underscoring the role of programming in enhancing understanding of key AI concepts such as algorithms and technical applications. However, the gap in AI ethics scores ( $4.38 \pm 0.54$  vs.  $4.33 \pm 0.58$ ) was not statistically significant ( $p = 0.085$ ).

Overall, students with programming experience demonstrated superior AI literacy, particularly in the AI knowledge and AI capability. Moving forward, it is crucial to extend and enhance programming and other technical skill-building courses, strengthen students' technical foundations and foster their comprehensive AI literacy, particularly in AI knowledge and AI capability.

### 5.4.5 Differences in Participation in Digital Intelligence Education

[Figure 9](#) outlines the differences in AI literacy scores among undergraduates based on their involvement in Wuhan University's core courses on digital intelligence education. Using an independent-samples *t*-test, the analysis revealed that students who participated in the core courses on digital intelligence scored significantly higher in AI attitude, AI knowledge, AI capability, and total scores compared to those who did not participate. However, the differences in AI ethics scores were not statistically significant ( $p = 0.064$ ).



**Figure 9** Differences in AI literacy scores for participants of core digital intelligence education courses at Wuhan University.

Specifically, participants in the core courses on digital intelligence scored significantly higher in AI knowledge ( $3.82 \pm 0.65$ ) compared to non-participants ( $3.69 \pm 0.68$ ), marking the most significant difference ( $p = 0.000$ ). This suggests that engagement in these core courses significantly enhances students' understanding of AI fundamentals, technical applications, and cutting-edge developments. Additionally, compared with non-participants, participants demonstrated notable

advantages in AI attitude ( $4.19 \pm 0.65$  vs.  $4.12 \pm 0.73$ ) and AI capability ( $4.01 \pm 0.61$  vs.  $3.94 \pm 0.65$ ), likely reflecting the courses' positive impact on attitude development and practical skills improvement (see [Figure 9](#)). However, in AI ethics ( $4.38 \pm 0.54$  vs.  $4.33 \pm 0.58$ ,  $p = 0.064$ ), while participants scored slightly higher, the difference was not statistically significant.

#### 5.4.6 Comparison of Self-Assessment Levels and Actual Scores

[Table 20](#) illustrates the relationship between participants' self-assessed AI literacy levels (ranging from "no understanding at all" to "proficient") and their actual measured scores. The results show a strong correlation, with average scores increasing as self-assessment levels improve, indicating that participants' self-perception aligns closely with their actual performance. Specifically, participants who rated themselves as having "no understanding at all" had an average score of 3.46, while those who self-assessed as "proficient" achieved an average score of 4.84. Notably, the most significant improvement occurred between the "somewhat proficient" and "proficient" levels, with a score increase of approximately 0.40 points. This finding highlights the accuracy of students' self-perceptions regarding their AI knowledge and AI capability, providing the basis for the design of educational interventions. It also suggests that self-assessment can serve as an effective reference for evaluating AI literacy. Educators can utilize self-assessment data to identify skill gaps quickly and implement tailored teaching strategies to enhance comprehensive AI literacy of students across different competency levels.

**Table 20** Self-assessment score

Category	Average score
No understanding at all	3.46
Basic understanding	4.00
Familiar	4.29
Somewhat proficient	4.44
Proficient	4.84

## 6 Discussion

The AILES-CS developed in this study establishes a scientific foundation for systematically assessing college students' AI literacy by defining four core dimensions: AI attitudes, AI knowledge, AI capability, and AI ethics. Furthermore, the accompanying assessment scale enables quantitative diagnosis of competency gaps in

individuals or groups while identifying disparities influenced by disciplinary backgrounds and technical proficiency. This tool empowers universities to optimize curriculum design, implement targeted training programs, and advance AI education from "generalized promotion" to "precision cultivation." By bridging digital divides and addressing educational inequities, AILES-CS serves as an effective instrument to foster equitable AI competency development.

The empirical analysis in this study explored multiple dimensions of AI literacy among students at Wuhan University, including gender, academic year, academic background, technical foundation, and engagement in digital intelligence education. The results reveal the following key findings:

(1) Gender did not significantly affect AI literacy levels, while academic year and academic background had significant impacts on AI literacy levels.

(2) AI literacy scores were relatively high among first-year undergraduates, likely due to Wuhan University's recent introduction of digital intelligence courses (e.g., Introduction to Data Science, AI Fundamentals, and AI and Machine Learning). These courses, aligned with societal demands, have significantly enhanced students' AI attitudes, knowledge, and capabilities.

Moreover, the impact of disciplinary background on AI literacy is particularly significant. Students from the Faculty of Information Science excelled in AI knowledge and AI capability, likely due to the inclusion of more technical content in their curriculum. In contrast, students from the Faculty of Humanities and the Faculty of Social Sciences had comparatively weaker performance in technical dimensions of AI literacy. However, the rapid advancement of AI tools is exerting a profound influence on the humanities and social sciences, with applications already demonstrating unique functions and impacts in areas such as human behavior simulation ([Chang et al., 2024](#)), language learning ([Peng et al., 2023](#)), peer review and publishing ([Liu & Shah, 2023](#)), and speech recognition ([Latif et al., 2023](#)). To address this disparity, it is essential for non-technical disciplines to increase their focus on AI education. This could involve introducing courses such as basic programming and AI fundamentals into traditional non-technical curricula, as well as integrating AI-related educational elements into existing specialized courses. Such efforts would foster interdisciplinary educational integration, injecting new vitality into the comprehensive enhancement of students' AI literacy. Furthermore, students with technical backgrounds and those who participated in digital intelligence education

courses demonstrated significantly higher AI literacy levels, particularly in AI knowledge and AI capability. Programming experience, in particular, was associated with superior performance in the AI attitude, AI knowledge, and AI capability. However, differences in the AI ethics were not statistically significant, suggesting that AI ethics are more influenced by personal beliefs and societal values than by technical expertise.

Based on these findings, it is advisable for Wuhan University to further strengthen the practice and application of digital intelligence education by expanding the coverage of digital intelligence courses. Emphasis should be placed on combining AI knowledge, technical practice, and ethics education into a comprehensive and structured curriculum. To address varying disciplinary backgrounds and ability levels of students, a tiered teaching strategy should be implemented. This approach should promote interdisciplinary AI education, especially in traditionally non-technical fields such as the humanities, social sciences, and medicine. Key measures could include increasing the penetration of technical content and practical education related to AI in these disciplines, along with adding foundational programming and AI introductory courses. Additionally, Wuhan University could establish collaboration with enterprises to link digital intelligence education with real-world application scenarios. Providing more hands-on opportunities for students would further enhance their comprehensive skills in technical applications, equipping them for success in the digital intelligence era. This would ensure that digital intelligence education serves as a cornerstone of talent cultivation at Wuhan University.

At the state level, empowering higher education in China with AI requires enhanced top-level design to establish a national framework for AI education. This includes optimizing curricula and promoting interdisciplinary integration to foster the systematic and standardized development of AI literacy education. The aim is to comprehensively enhance students' understanding and application of AI technologies. Advanced AI technologies should also be leveraged to build intelligent learning support platforms. Examples include Wuhan University's Digital Intelligence Education Practice and Innovation Platform and the LuoJia Online AI Intelligent Teaching Center. These platforms can facilitate personalized education, promote educational equity, and enable resource sharing. Moreover, scientific evaluation mechanisms and tools such as AILES-CS should be developed to dynamically optimize course content and teaching methods. This would ensure continuous improvement in the quality of AI education, laying a solid foundation

for cultivating well-rounded talent with technical expertise, social responsibility, and innovative spirit.

On a theoretical level, this research refines the theoretical framework for AI literacy and expands the applicability of existing evaluation systems. It provides academic support for defining the connotations of AI literacy and developing assessment methods. On a practical level, the findings offer useful insights for higher education reform and curriculum design. They are particularly meaningful for enhancing students' AI literacy in the context of digital intelligence education. The study validates the effectiveness of Wuhan University's digital intelligence education for undergraduates, highlighting its positive impact on fostering students' AI overall literacy in knowledge, capability, and other dimensions. Furthermore, the results provide a practical reference for other higher learning institutions seeking to optimize their AI education systems in the digital intelligence era.

## 7 Limitations

First, this study focused exclusively on undergraduate students at Wuhan University, which may limit the generalizability of the findings. Future research will leverage the Digital Intelligence Education Practice and Innovation Alliance to expand the sample size by incorporating higher learning institutions of various types and from different regions, and to further validate the broader applicability of the results. Second, the assessment criteria used in this study primarily emphasized static assessments, without fully accounting for dynamic and contextual factors. This limitation constrains the measurement of AI literacy in real-world application scenarios. Future work will optimize the assessment criteria using tools such as the AILES-CS framework, incorporating contextualized designs and experimental methods to compare students' AI literacy across diverse scenarios. Additionally, this study did not deeply investigate the effects of different educational models, such as traditional classroom-based teaching versus practice-oriented instruction, on AI literacy. Future research could conduct comparative studies of these pedagogical approaches to provide more practical guidance and theoretical support for enhancing AI literacy in higher education.

## 8 Conclusions

This study provides a comprehensive analysis and summary of AI literacy among students at Wuhan University from two perspectives: metric development and empirical investigation. On metric development,

the study utilized the KSAVE model and UNESCO (2025a) framework to establish AILES-CS, an evaluation system comprising four primary dimensions: AI attitude, AI knowledge, AI capability, and AI ethics. The Delphi method and analytic hierarchy process were employed to define the specific content and weight distribution of each dimension. Based on this framework, a 31-item AI literacy assessment scale was developed. Reliability and validity testing demonstrated the scale's strong measurement stability and structural consistency, offering a robust tool for scientifically assessing students' AI literacy levels.

On empirical research, the study collected data from 1,651 undergraduate students at Wuhan University through a questionnaire survey. The results revealed that students exhibited overall high levels of AI literacy, with positive attitudes and skills. However, the analysis also identified discrepancies across dimensions, particularly in AI knowledge acquisition, where notable gaps were observed. Further analysis showed that gender had no significant impact on AI literacy, whereas factors such as technological background and participation in digital intelligence education courses significantly enhanced students' AI knowledge and AI capability. In particular, programming experience and involvement in core digital intelligence courses were found to markedly improve students' technical understanding and practical skills. The study's findings provide practical insights for optimizing undergraduate education at Wuhan University and serve as a reference for advancing AI literacy education in higher education institutions.

**Acknowledgments** This study was funded by the Hubei Provincial Undergraduate Higher Education Provincial-Level Teaching Reform Research Project "Exploration and Practice of the 'Four-in-One' Digital Intelligence Education Transformation Path in Higher Education Institutions" and Wuhan University Undergraduate Education Quality Construction Comprehensive Reform Project Digital Intelligence Field Research Special Topic "Evaluation of Artificial Intelligence Literacy and Construction of Professional Map in Universities in the Digital Intelligence Era."

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Ethics Statement** The authors declare that their Institutional Ethics Committee confirmed that no ethical review was required for this study. Written informed consent for participation was not required because all participants' data was anonymized before the statistical analyses were done.

**Data Availability Statements** The authors confirm that all data generated or analysed during this study are included in this published article.

## References

- Andresen, S. L. (2002). John McCarthy: Father of AI. *IEEE Intelligent Systems*, 17(5), 84–85.
- Agre, P. E. (1972). *What to read: A biased guide to AI literacy for the beginner*. Cambridge: MIT Artificial Intelligence Laboratory.
- Bewersdorff, A., Hornberger, M., Nerdel, C., & Schiff, D. (2024). AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy. *Computers and Education: Artificial Intelligence*, 8, 100340.
- Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10(1), 29.
- Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, 107468.
- Černý, M. (2024). University students' conceptualisation of AI literacy: Theory and empirical evidence. *Social Sciences*, 13(3), 129.
- Chang, Y., Wang, X., Wang, J., Yuan, W., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P.S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 39.
- Chiu, T. (2023). The impact of generative AI (genAI) on practices, policies and research direction in education: A case of ChatGPT and midjourney. *Interactive Learning Environments*, 32(10), 1–17.
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google, is it ok if I eat you?" Initial explorations in child-agent interaction. In: *Proceedings of the 2017 Conference on Interaction Design and Children, Stanford*. New York: ACM, 595–600.
- Gill, K. S. (1991). Artificial intelligence for social citizenship: Toward an anthropocentric technology. *Applied Artificial Intelligence*, 5(1), 15–27.
- Gratch, J., Lucas, G. M., King, A. A., & Morency, L. P. (2014). It's only a computer: The impact of human-agent interaction in clinical interviews. In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, Paris*. Richland: International Foundation for Autonomous Agents and Multiagent Systems, 85–92.
- Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer.
- Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S. & Huber, P. (2016). Artificial intelligence and computer science in education: From kindergarten to university. In: *Proceedings of the 2016 IEEE Frontiers in Education Conference*. Erie: IEEE, 1–9.
- Karaca, O., Çalışkan, S. A., & Demir, K. (2021). Medical artificial intelligence readiness scale for medical students (MAIRS-MS)—Development, validity and reliability study. *BMC*

- Medical Education*, 21, 1–9.
- Latif, S., Usama, M., Malik, M. I., & Schuller, B. W. (2023). Can large language models aid in annotating speech emotional data? Uncovering new frontiers. *arXiv Preprint*, arXiv:2307.06090.
- Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning*, 9(1), 50.
- Liu, R., & Shah, N. B. (2023). ReviewerGPT? An exploratory study on using large language models for paper reviewing. *arXiv Preprint*, arXiv:2306.00622.
- Ministry of Education of the People's Republic of China (MoE). (2024a). *Notice from the Ministry of Education on Publishing "AI Innovation Action Plan for Institutions of Higher Education"*. Available from Ministry of Education website.
- Ministry of Education of the People's Republic of China (MoE). (2024b). *Notice from the Department of Higher Education, Ministry of Education, on Publishing the First Batch of Typical Cases of "Artificial Intelligence + Higher Education" Application Scenarios*. Available from Ministry of Education website. (in Chinese)
- 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(1), 100041.
- Partridge, D. (1987). The scope and limitations of first generation expert systems. *Future Generation Computer Systems*, 3(1), 1–10.
- Peng, L., Nuchged, B., & Gao, Y. (2023). Spoken language intelligence of large language models for language learning. *arXiv Preprint*, arXiv:2308.14536.
- Pereira, A., Prada, R., & Paiva, A. (2014). Improving social presence in human-agent interaction. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto*. New York: ACM, 1449–1458.
- Scammell, A. (2000). Visions of the information future. *Aslib Proceedings*, 52(7), 264–269.
- Su, G. (2018). Unemployment in the AI age. *AI Matters*, 3(4), 35–43.
- Su, W., Guo, H., Lu, Z., Pan Y., & Liu G. (2024). Construction of artificial intelligence literacy evaluation index system and validation of its effectiveness in China's college and university student population. *Library Development*, 1–25. (in press) (in Chinese).
- Tenório, K., Olari, V., Chikobava, M., & Romeike, R. (2023). Artificial intelligence literacy research field: A bibliometric analysis from 1989 to 2021. In: *Proceedings of the 54th ACM Technical Symposium on Computer Science Education, Toronto*. New York: ACM, 1083–1089.
- UNESCO. (2025a). AI competency framework for students. Available from UNESCO website.
- UNESCO. (2025b). K-12 AI curricula: A mapping of government-endorsed AI curricula. Available from UNESCO website.
- 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.
- Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, 29(4), 4785–4808.
- Wu, D. (2024). *Wuhan University AI literacy evaluation guide*. Wuhan: Wuhan University Press.