Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Fragmentation to Integration: A Delphi-Validated AI Literacy Competency Framework for Higher Education Teachers in China

Muhammad Arif
Southwest University
Miracle Uzochukwu Okafor
Hanjiang Normal University
Mirajur Rhaman Shaoan
Southwest University
Wang Yong
wangycn@swu.edu.cn
Southwest University
Bushra Jamil
Southwest University

Research Article

Keywords: artificial intelligence literacy, higher education, teacher competence, Delphi method, AI-TPACK, China

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10634652/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence (AI) is transforming higher education, yet frameworks for developing teachers' AI literacy in China remain fragmented across policy documents, institutional initiatives, and theoretical perspectives. This study developed and established expert consensus on the content validity of an integrated AI literacy competency framework for higher education teachers. Employing an exploratory sequential mixed-methods design, the study was conducted in two phases. Phase 1 involved a PRISMA-guided comparative document analysis of four Chinese policy and institutional frameworks, supplemented by two theoretically derived comparator models. These frameworks were examined across five analytical dimensions, revealing three central tensions: standardisation versus contextualisation, static competency structures versus developmental progression, and instrumental versus agentic conceptualisations of AI. Based on these findings, a preliminary three-layer framework comprising 27 indicators was developed. Phase 2 applied a two-round Delphi process with 15 scholars and higher education administrators to evaluate and refine the proposed framework. Through expert feedback, one indicator was removed, two indicators were merged, three new indicators were incorporated, and Comprehensive Ethics was repositioned as a cross-cutting dimension. The final framework retained 27 indicators organised across three developmental layers: Foundation, Practice, and Innovation. Twenty-six indicators (96.3%) achieved the predefined consensus criteria, forming the consensus-supported core, while L2.7 (Classroom Dynamics with AI) was retained as a provisional indicator requiring further investigation. The framework extends AI-TPACK by incorporating developmental trajectories, contextual adaptation, and ethical integration. It offers an evidence-informed foundation for institutional implementation, teacher professional development, policy formulation, and future assessment development. Further psychometric testing and classroom-based validation are required to establish its measurement properties and practical applicability.

1. Introduction

1.1 The Global Imperative: AI in Higher Education

Generative AI has entered higher education at remarkable speed. Large language models such as ChatGPT and DeepSeek can draft text, solve structured problems, generate code, and provide conversational support across specialised subjects. These capabilities create opportunities for personalised learning, formative feedback, assessment support, intelligent tutoring, and more efficient research workflows (Celik, 2023; Crompton & Burke, 2023; UNESCO, 2023). They also intensify concerns about academic integrity, data privacy, algorithmic bias, reliability, and the changing meaning of learning and authorship (Akgun & Greenhow, 2022; UNESCO, 2023).

Teachers mediate between technological capability and educational practice. Accordingly, AI literacy requires more than operational familiarity with tools; it involves understanding AI concepts and limitations, evaluating outputs critically, applying AI pedagogically, and acting ethically in context (Long & Magerko, 2020; Ng et al., 2021; UNESCO, 2024). Governments and educational institutions have therefore begun to formulate standards and frameworks intended to guide teachers' professional learning and responsible AI use.

1.2 The Chinese Policy Context

China has made artificial intelligence and educational digitalisation national priorities. The Next Generation Artificial Intelligence Development Plan (State Council, 2017) set the goal of China becoming a global AI leader by 2030 and made education a priority. Successive strategies, including the Education Modernisation 2035 (Ministry of Education, 2019) and the Plan for the Construction of a Digital China (Communist Party of China Central Committee & State Council, 2023), connected digital transformation and educational reform.

These strategies have become more specific and concrete. The Education Powerhouse Construction Plan (2024–2035) published in 2025, asked for the national education digitalisation strategy to be advanced, and for transformation of education with support of artificial intelligence (Communist Party of China Central Committee & State Council, 2025). In

2026, five national departments published the "AI + Education" Action Plan, together establishing coordination in the area of AI literacy and talent, as well as educational applications, infrastructure, and governance (Ministry of Education et al., 2026).

An essential basis is provided by the JY/T 0646–2022 Teacher Digital Literacy standard which divides Digital Literacy into five dimensions: Digital Awareness, Digital Technologies Knowledge and Skills, Digital Applications, Digital Social Responsibility, and Professional Development (Ministry of Education, 2022). This national standard, however, only addresses Digital Literacy, and cannot be classified as a generative-AI specific competency framework. In the absence of a more AI-focused framework, provincial and institutional documents have provided more specific AI-based approaches. These documents differ in structure, terminology, and logic of development, as well as the intended purpose and use. The Interim Measures for the Management of Generative Artificial Intelligence Services, co-issued by seven national departments, offer a broader, regulatory framework for the responsible use of AI (Cyberspace Administration of China et al., 2023).

1.3 The Research Gap

Three interrelated gaps motivate this study. First, the literature search identified no study that had systematically compared the selected Chinese policy and institutional frameworks using a common analytical scheme encompassing their theoretical foundations, structural logic, conceptualisations of AI, assessment orientations, and levels of specificity. Second, the search did not identify an AI-specific competency framework for higher education teachers in China that had undergone a transparent Delphi process using prespecified consensus criteria. Third, the literature lacks a consensus-supported and developmentally organised framework that can guide institutional piloting, professional learning, and the subsequent development and psychometric validation of assessment instruments. These are gaps in comparative synthesis and content validation, not evidence that the existing policy standards lack authority or practical value.

1.4 Research Purpose and Questions

This study employed an exploratory sequential mixed-methods design. Phase 1 used systematic comparative document analysis to examine four Chinese documentary frameworks alongside two separately constructed theoretical comparators (Bowen, 2009). In Phase 2 the two-round Delphi method (Hasson et al., 2000; Keeney et al., 2011) was used to assess and improve the first version of the meta-framework.The study addressed four research questions.

The study was guided by four research questions: (1) What theoretical foundations, dimensional structures, and developmental logics characterise the selected Chinese higher education teacher AI literacy frameworks?; (2) How do the selected frameworks converge and diverge in their conceptualisation of AI's role (tool, medium, or co-agent) and their assessment orientation (diagnostic, formative, or summative)?; (3) What preliminary meta-framework can be synthesised from the comparative analysis to address the identified tensions?; and (4) To what extent do educational experts support the proposed meta-framework's content, structure, and developmental logic?

1.5 Significance of the Study

Four contributions can be identified in this study. In theory, this study expands the components of developmental progression, contextual adaptability, and ethics within the layers of the framework of AI-TPACK. In terms of methodology, this study integrates both a comparative document analysis and a Delphi technique. Practically, it provides a consensus-supported foundation that can be piloted by policymakers, university leaders, and teacher educators before being translated into a measurement instrument. Internationally, it contributes evidence from a centrally governed, non-Western higher education context to discussions of AI competency modelling (Eyal, 2025; UNESCO, 2024).

2. Literature Review and Theoretical Framework

2.1 International Frameworks for Teacher Digital and AI Competence

Three international frameworks provide the principal reference points for this study. DigCompEdu (Redecker, 2017) presents six aspects for conceptualising the digital competence of teachers and defines both the Teaching and Learning and the

Assessment aspects along with the Empowering Learners and the Facilitating Learners' Digital Competence aspects. Each of the six aspects has six levels of proficiency. Since DigCompEdu is prior to the expansion of generative AI, it lacks the specification of teacher competencies related to AI.

Teaching in the Age of AI (Williamson, 2023) and the UNESCO AI Competency Framework for Teachers (UNESCO, 2024) place AI in its central and definitional place. The UNESCO AI Competency Framework for Teachers arranges 15 competencies within five domains and three levels of progression (Acquire, Deepen, Create) and foregrounds ethical and responsible use of AI, engagement with the AI basics, and teaching with AI (UNESCO, 2024).

The combination of TPACK and the models of AI-TPACK offers a further dimension in terms of pedagogy (Celik, 2023; Mishra & Koehler, 2006; Ning et al., 2024). Central to AI-TPACK models is the consideration of AI-related technologies, their capabilities and limitations, and the pedagogical affordances and the interplay with pedagogical and content knowledge. Unlike the other models, AI-TPACK has primarily considered the relationships of the knowledge spheres and not a pathway of staged or graduated AI literacy for teachers.

While these models are International in scope and therefore necessarily broad, focusing them on a specific national context will require consideration of local policy, fields of practice, the frameworks of governance, and the professional context (Eyal, 2025; UNESCO, 2024).

2.2 The Evolution of Chinese Teacher Digital Literacy Policy

For analytical purposes, we organise the evolution of Chinese teacher digital literacy policy into three periods. This periodisation is an author-developed interpretive synthesis rather than an official government classification. Before 2020, policy attention centred primarily on information and communication technology application and teachers' operational capacity. From 2020 to 2023, the focus started to widen to digital literacy, including cognitive, ethical, professional, and applied dimensions, culminating into the Teacher Digital Literacy standard (Ministry of Education, 2019, 2022).

From 2024, generative AI and AI literacy started to attract more attention in educational policy and practice. More provincial and university initiatives started to develop AI-specific requirements. At the same time, national policy started addressing teaching, talent development, AI, teaching infrastructure, and governance of AI (Ministry of Education et al., 2026; UNESCO, 2024). This progression demonstrates responsiveness to technological change, but it has also produced a recent and heterogeneous collection of AI-related competency formulations.

2.3 Theoretical Lenses for Framework Analysis

Four theoretical perspectives informed the analytical framework. First, teacher development theory provided a developmental lens. Fuller's (1969) concerns model describes movement from self-related concerns to task-related and impact-related concerns, whereas Huberman (1989) presents a more complex professional career cycle. We therefore used these accounts as complementary developmental lenses, not as a single three-stage scale.

Second, AI-TPACK placed a pedagogical lens on the integration of AI into pedagogical and disciplinary contexts. This lens goes beyond the conceptualization of AI integration as a technical skill (Celik, 2023; Mishra & Koehler, 2006; Ning et al., 2024). Third, institutional theory focused on the coercive, mimetic, and normative pressures of the adoption and standardization of competency frameworks (DiMaggio & Powell, 1983; Meyer & Rowan, 1977).

Fourth, the activity theory and the distributed cognition theory offered related but differentiated views in sociocultural theory. Activity theory centers on systems of activity that are mediated, their rules, communities, divisions of labor, and the internal contradictions of the system (Engestrom, 1987, 2001). Distributed cognition considers the distribution of cognitive activity over people, artifacts, and the environment (Salomon, 1993). These theories facilitated the understanding of whether AI was conceptualized as a tool with a pedagogical function, a medium that acts to integrate pedagogical activities, or as a coupled agent in the teaching and learning process. The categories presented in this study are an analytical synthesis, not a sequential phase or step, of the theorists cited.

2.4 Gaps in the Literature

The reviewed literature revealed two linked gaps. First, we found no published comparison of the selected Chinese frameworks across national, provincial, and institutional levels using a common analytical scheme. Second, we found no published AI-specific competency framework for higher education teachers in China that combined comparative synthesis with transparent Delphi-based content validation. Addressing these gaps requires distinguishing existing documentary frameworks from author-constructed theoretical comparators and distinguishing expert consensus from subsequent psychometric or predictive validation (Bowen, 2009; Hasson et al., 2000; Keeney et al., 2011).

3. Methodology

3.1 Research Design: Exploratory Sequential Mixed Methods

We employed an exploratory sequential mixed-methods design in which the qualitative findings from Phase 1 informed the development of the preliminary meta-framework evaluated in Phase 2. Phase 1 comprised a systematic comparative document analysis that identified cross-framework patterns and supported the synthesis of the preliminary meta-framework. Phase 2 combined quantitative ratings and qualitative comments within a two-round Delphi process to evaluate and refine the framework's content and structure. Table 1 summarises the methods and purposes of the two phases.

Table 1
Overview of the Sequential Mixed-Methods Design

Phase	Method	Purpose
Phase 1	Qualitative comparative document analysis	Compare four Chinese documentary frameworks and two separately constructed theoretical comparators; synthesise a preliminary meta-framework.
Phase 2	Two-round Delphi: quantitative ratings and qualitative comments	Evaluate, revise, and establish expert consensus concerning the preliminary framework's indicators and structural logic.

3.2 Phase 1: Comparative Document Analysis

3.2.1 Source Selection and PRISMA-Guided Search

A systematic search was conducted from 15 January to 15 March 2026. The academic databases searched were CNKI, Wanfang, CQVIP, Web of Science, Scopus, and Google Scholar. Additional searches covered Ministry of Education webpages, provincial education department portals, university websites, reference lists, and related policy sources. Search strings, source-specific adaptations, search dates, and screening decisions are documented in Appendix A.

Database searches yielded 650 records (CNKI, 180; Wanfang, 150; CQVIP, 120; Web of Science, 80; Scopus, 65; and Google Scholar, 55). Policy-website searches and snowballing yielded 50 additional records, producing 700 records in total. After 120 duplicates were removed, 580 records were screened. Of these, 540 were excluded during title and abstract or executive-summary screening, and 40 reports underwent full-text assessment. The documentary search retained four eligible Chinese frameworks. Two author-synthesised theoretical comparators were then added separately as analytical benchmarks; they were not counted as documentary records selected through the PRISMA process.

Documentary sources were included in cases when they (a) were published by an official national, provincial, or higher education body; (b) were targeted toward higher education teachers; (c) were focused on artificial intelligence competence or, regarding the national Teacher Digital Literacy standard, were focused on the relevant digital-literacy baseline; (d) were published in the time span between 2022 and 2026; (e) were sufficiently detailed for coding. Sources were excluded when they were focused on K-12 or vocational education, did not include relevant artificial intelligence or digital literacy content, were a duplicate of a source and/or provided insufficient data for analysis.

The theoretical comparators were constructed through a separate synthesis of peer-reviewed and authoritative international literature. They were included to test alternative pedagogical, developmental, and role-based logics rather than to represent additional Chinese policy documents. The analytical corpus comprised four documentary frameworks and two author-synthesised comparators. The final analytical corpus is presented in Table 2.

Table 2
Final Analytical Corpus

Code	Source or analytical unit	Type	Year
F1	Ministry of Education Teacher Digital Literacy Standard	National policy standard	2022
F2	Zhejiang provincial AI literacy framework for higher education teachers	Provincial framework	2024
F3	Tongji University teacher AI literacy standard	Institutional framework	2025
F4	Sichuan University staged teacher AI capacity framework	Institutional framework	2024
F5	AI-TPACK comparator, synthesised from Mishra and Koehler (2006), Celik (2023), and Ning et al. (2024)	Author-synthesised comparator	-
F6	Dual-track AI literacy comparator, synthesised from Eyal (2025) and UNESCO (2024)	Author-synthesised comparator	-

Note. F5 and F6 were analytical comparators rather than independently issued policy or institutional frameworks. Appendix A reports the official Chinese titles, verified English translations, issuing bodies, stable source locations, and access dates for F1-F4.

3.2.2 Analytical Coding Scheme

Five analytical dimensions were developed iteratively through team discussion and applied to all six framework/model units (Bowen, 2009). The framework-by-dimension judgement was the primary coding unit. Two researchers independently coded all 30 judgements (six units by five dimensions) before reconciliation. The complete pre-reconciliation dataset yielded 87% agreement and Cohen's kappa of .82. Disagreements were resolved through discussion, with third-researcher adjudication when required. The coding book recorded source excerpts and locations, initial and reconciled codes, and decision rationales. Table 3 summarises the dimensions, analytical questions, and principal coding categories.

Table 3
Analytical Coding Scheme

Dimension	Analytical question	Principal coding categories
D1: Theoretical grounding	What theoretical or policy logic primarily drives the unit?	Policy-driven; pedagogy-driven; technology-driven; ecosystemic or relational.
D2: Structural logic	How is competence organised?	Static list; developmental stages; matrix-based structure.
D3: AI role	How is AI conceptualised in relation to teaching and learning?	Bounded tool; independent pillar; integration medium; co-agent.
D4: Assessment orientation	What assessment or developmental purpose is emphasised?	Diagnostic; formative; summative; developmental.
D5: Target granularity	For whom and at what level is the unit intended?	Universal; discipline-differentiated; stage-differentiated; institution-specific.

3.2.3 Data Analysis Procedure

Analysis proceeded in three stages. First, within-case analysis produced a structured profile of each framework/model unit across the five dimensions. Second, cross-case comparison identified areas of convergence, divergence, and conceptual

tension. Third, thematic aggregation was used to synthesise the preliminary three-layer meta-framework. The preliminary indicators and their source-to-framework mappings were retained in an audit trail reported in the supplementary materials.

3.3 Phase 2: Delphi Expert Validation

3.3.1 Expert Panel Selection

Experts were recruited purposively to provide relevant academic, administrative, disciplinary, and policy experience. Fifteen of the eighteen experts that were invited to take part responded to the questionnaire (83.3% response rate). They came from eight different provinces across China. To qualify, people needed (a) a doctorate in education, educational technology, or a related field, (b) a minimum of 10 years of work experience in a tertiary education institution, (c) at least 3 peer-reviewed articles on digital or AI literacy, and (d) a senior academic or administrative post. Their curriculum vitae were reviewed to confirm eligibility. There were no conflicts of interest, and participation was voluntary and not paid.

3.3.2 Anonymity and Ethical Protocol

Delphi anonymity was preserved in the study (Keeney et al., 2011). In the rounds, the panel did not know the identities or affiliations of other participants, nor the identity of the respondents. They only received feedback in an aggregated form, and only grouped demographic data were shared. Informed consent was obtained from all participants. The Ethics Review Board, School of History and Culture, Southwest University, provided ethical approval (HCS-ERB-2026-089; 26 November 2025).

3.3.3 Delphi Implementation and Consensus Criteria

The protocol allowed for a maximum of two Delphi rounds. In Round 1, experts evaluated 27 preliminary indicators for their importance and clarity, each using a five-point scale, and provided open-ended comments. The aggregate results of Round 1 and a synthesis of the qualitative feedback formed the basis of Round 2. In Round 2, experts had the opportunity to rate again eight indicators that had been formulated in Directly Comparable terms to the Round 1 formulation. They also evaluated a merged formulation of L3.6 and L3.8, rated four proposed additions (N1-N4), assessed the proposed output of a cross-cutting merger (N5), and reviewed the other structural propositions that were provided. Table 4 provides a summary of the activities and timeframes of each Delphi round.

Table 4
Delphi Round Structure

Round	Activities	Timing
Round 1	Rated 27 preliminary indicators for importance and clarity; supplied open-ended comments.	Week 1
Round 2	Reviewed aggregated feedback; re-rated eight directly comparable indicators; evaluated one merged indicator, N1-N5 outputs, and structural proposals.	Week 3

Consensus required the simultaneous attainment of all three prespecified criteria: a mean importance rating of at least 4.0, a standard deviation no greater than 1.0, and at least 80% of experts assigning a rating of 4 or 5 (Hasson et al., 2000; Keeney et al., 2011). Clarity ratings informed wording revisions but did not determine consensus. Because the protocol limited the study to two rounds, an indicator remaining below the thresholds after Round 2 was designated provisional or rejected according to the documented panel feedback and structural decision.

3.3.4 Data Analysis

The means (M), standard deviations (SD), and the percentage of experts who rated a given item a 4 or 5 were used to assess ratings of importance and clarity. Consensus on importance ratings was determined solely based on the three specified criteria. While clarity ratings led to revisions of the survey statements, they did not factor in the assessment of consensus. Thematic analyses of the qualitative comments focused on described concerns regarding redundancy, omissions, wording, and the structure of the framework (Braun & Clarke, 2006).

Changes made between survey rounds were examined for the eight indicators that had essentially the same description in Round 1 and Round 2. The analyses involved two-sided Wilcoxon signed-rank tests. The Pratt method was used for cases of zero difference, and for cases of tied absolute differences, average ranks were computed. All 15 experts participated in both rounds of the survey, resulting in a total of 15 paired cases of survey responses for each indicator. Based on the work of Rosenthal (1994), the effect size was computed and the following was produced.:

$$r = \frac{|Z|}{\sqrt{N}}$$

1

where (|Z|) represents the absolute value of the standardised Wilcoxon test statistic and *N* represents the number of paired observations. The current analysis includes N = 15. As the comparison between rounds was exploratory, the p values reported were not adjusted. Thus, p values were interpreted as descriptive evidence of rating change. The presence of statistical significance or the magnitude of the effect size were not considered in the evaluation of consensus.

4. Findings

4.1 Phase 1: Comparative Document Analysis

4.1.1 Comparative Patterns Across the Framework/Model Units

The comparative document analysis examined six framework/model units: four Chinese policy or institutional frameworks (F1-F4) and two author-synthesised theoretical comparators (F5-F6). The two source types were retained as analytically distinct categories. The policy and institutional documents represented existing standards or organisational approaches, whereas the theoretical comparators were constructed to test alternative pedagogical and developmental logics. Each unit was coded across five dimensions: theoretical grounding, structural logic, conceptualisation of AI's role, assessment orientation, and target granularity.

The analysis revealed substantial structural and conceptual variation (Table 5). F1 and F2 emphasised broad standardisation and policy applicability, whereas F3 offered a more relational account of AI and F4 provided the clearest staged progression. The AI-TPACK comparator foregrounded the interaction of technological, pedagogical, and disciplinary knowledge but did not specify a developmental sequence. The dual-track comparator differentiated general and discipline-specific competence while also incorporating progression from foundation to leadership. These contrasts provided the analytical basis for the three cross-framework tensions reported below.

Table 5

Comparative Matrix of the Six Framework/Model Units

Source	Theoretical grounding	Structural logic	AI role	Assessment orientation	Target granularity
F1: National (2022)	Policy-driven	Static list	Sub-set tool	Summative/diagnostic	National/general
F2: Zhejiang (2024)	Policy and scenario based	Static 3 x 3 matrix	Independent pillar	Diagnostic	Province-wide
F3: Tongji (2025)	Ecosystemic/relational	Static list	Community member/co-agent	Formative	Institution-wide; research-intensive context
F4: Sichuan (2024)	Stage-oriented	Dynamic four-stage pathway	Evolving tool	Formative/developmental	Institution-wide
F5: AI-TPACK	Pedagogy-driven	Interwoven knowledge matrix	Integration medium	Diagnostic	Discipline differentiated
F6: Dual-track comparator	Differentiation and progression	Dual-track developmental matrix	Role varies by track	Formative/developmental	General and specialised tracks

Note. F1 = national Teacher Digital Literacy Standard; F2 = Zhejiang provincial framework; F3 = Tongji University framework; F4 = Sichuan University framework; F5 and F6 = author-synthesised theoretical comparators. Classifications of structural logic, AI role, assessment orientation, and target granularity represent the authors' analytical coding.

4.1.2 Three Cross-Framework Tensions

Three tensions recurred across the six units. The first concerned standardisation versus contextualisation. F1 and F2 prioritised broadly applicable expectations, while F5 and F6 made disciplinary or role-based differentiation more explicit. The second concerned static versus developmental organisation. Most F1, F2, F3, and F5, structured competence in lists or matrices, while F4 and F6, used staged progression. The third, focused on instrumental versus agentic views of AI. F1, F2, and F4, mainly, viewed AI as a bounded or developing tool. F5, placed AI in the framework as a technology-pedagogy-content knowledge integration tool, and F3, had the most relational or agentic view. These results, indicated that the framework would require the integration of a common structure with the flexibility to incorporate staged progression, as well as a differentiated view of the relation of teachers with AI.

4.1.3 Synthesised Preliminary Meta-Framework

The initial version of the meta-framework comprised three layers: Foundation, Practice, and Innovation. This version was informed by the comparative synthesis and was described in the Delphi Study. Foundation comprised: Awareness of AI, AI knowledge and skills, knowledge of AI tools, and AI-related ethics; Practice comprised: the design of learning experiences and assessment, the teaching of learning experiences and assessment, and the context-sensitive integration of AI; and Innovation comprised: the design of curricula and programs, AI-informed research, and the professional activities and the role of leadership in the (re)design of programs and (re)structuring of organizations. Comprehensive Ethics was initially represented in the layers through indicators, and was reconsidered in the Delphi Study. The initial version of the framework described 27 indicators, allocated to each of the three layers in sets of nine. The generation of indicators and the related audit trail are reported in Appendix C.

4.2 Phase 2: Delphi Expert Validation

4.2.1 Expert Panel Characteristics

The second round of the Delphi study saw the same cohort of 15 panel members, setting a fantastic precedent of 100% retention rate. The Delphi panel members brought skills and experience at the intersection of academia, administration, discipline, and policy. The cohort reported a mean experience of 17.2 years (standard deviation = 4.8); additionally, 5 of the experts had prior experience in policy development. The details of the experts' cohort is presented in Table 6.

Table 6
Expert Panel Characteristics

Characteristic	**Value**
Total participants	15
Gender	Male, 10 (66.7%); female, 5 (33.3%)
Years of professional experience	M = 17.2 years; SD = 4.8
Professional role/rank	Full professor, 7; associate professor, 5; senior administrator, 3
Institutional type	Research-intensive university, 6; comprehensive provincial university, 5; vocational institution, 4
Primary disciplinary area	Educational technology/teacher education, 6; STEM, 4; humanities/social sciences, 3; medicine/arts, 2
AI/digital-literacy publications during previous five years	M = 6.8; SD = 3.2
Policy-development experience	5 experts (33.3%)

4.2.2 Round 1 Results

Consensus denoted the simultaneous satisfaction of three conditions: the average overall importance was ≥ 4.0, the overall standard deviation was ≤ 1.0, and 4 or 5 ratings were assigned to the indicators by ≥ 80% of the experts. In Round 1, 16 of 27 preliminary indicators (59.3%) attained consensus. Remaining were L1.2, L1.4, L1.8, L2.1, L2.5, L2.6, L2.7, L3.1, L3.3, L3.6, and L3.8. The remaining 11 indicators were to be addressed in Round 2.

The comments included four main changes: the alteration or integration of duplicate indicators, the inclusion of additional competencies, changes to the language of the indicators, and the endorsement of the three tiers combined with ethics throughout the entire framework. These comments informed the revised indicators and structural proposals presented to the panel in Round 2.

4.2.3 Round 2 Results

All 15 experts participated in Round 2. Consensus was reached on seven of the indicators that were re-rated in their original forms. L2.7 was still below the defined thresholds (M = 3.9, SD = 1.1; 73% agreement). L3.6 and L3.8 were merged and the newly constructed version reached consensus (M = 4.2, SD = 0.8; 87% agreement). Among the four proposed additions, consensus was reached for three: N1, N2 and N4. However, consensus was not achieved on N3. N5, the proposed cross-cutting outcome of merging L1.6 and L1.9, achieved consensus. Table 7 displays the Round 2 indicator and structural outcomes.

Table 7

Round 2 Indicator and Structural Outcomes

Item or decision	Round 1 M (SD)	Round 2 M (SD)	Round 2 agreement	Outcome
Panel A. Re-rated or merged indicators				
L1.2	3.8 (1.2)	4.2 (0.8)	87%	Consensus
L1.4	3.9 (1.1)	4.3 (0.7)	93%	Consensus
L2.1	3.7 (1.2)	4.1 (0.9)	80%	Consensus
L2.5	3.9 (1.0)	4.4 (0.7)	93%	Consensus
L2.6	4.1 (1.2)	4.3 (0.9)	87%	Consensus
L2.7	3.6 (1.3)	3.9 (1.1)	73%	No consensus; provisional
L3.1	3.8 (1.1)	4.0 (1.0)	80%	Consensus
L3.3	4.3 (1.2)	4.5 (0.8)	93%	Consensus
Merged L3.6/L3.8	-	4.2 (0.8)	87%	Consensus
Panel B. Candidate additions and cross-cutting merger output				
N1: AI-enhanced assessment design	-	4.8 (0.4)	100%	Accepted
N2: AI-enhanced accessibility	-	4.1 (0.9)	80%	Accepted
N3: Collaborative AI with students	-	3.8 (1.0)	73%	Rejected
N4: AI-assisted self-reflection	-	4.3 (0.7)	87%	Accepted
N5: Comprehensive Ethics	-	4.9 (0.3)	100%	Accepted as cross-cutting

Note. Consensus required M > = 4.0, SD < = 1.0, and at least 80% agreement. A dash indicates that no directly comparable Round 1 rating was available. Ninety-three percent of the panel supported merging L3.6 and L3.8; 87% subsequently rated the merged wording as 4 or 5. Eighty-seven percent supported removing L1.8.

Three structural changes were therefore implemented. First, L1.8 was removed. Second, L1.6 and L1.9 were merged into N5, Comprehensive Ethics, which was repositioned as a cross-cutting dimension. Third, L3.6 and L3.8 were merged into a single Innovation-layer indicator. N1, N2, and N4 were added, while N3 was rejected. These changes reduced duplication, strengthened the ethical architecture of the framework, and preserved the overall total of 27 indicators.

4.2.4 Between-Round Change

Exploratory Wilcoxon signed-rank tests were performed for the eight indicators that kept direct equivalents from Round 1. Out of these, the following indicators had statistically significant increases between rounds: L1.2, L1.4, L2.1, L2.5, L3.1, and L3.3. The indicators L2.6 and L2.7 had no statistically significant changes between the rounds. These tests assessed rating change rather than consensus; consequently, the non-significant result for L2.7 was not interpreted as evidence of stability or validity. Table 8 presents the complete exploratory between-round analysis.

Table 8

Exploratory Between-Round Change Analysis

Indicator	Z	p	Effect size r	Interpretation
L1.2	-2.81	.005	.73	Significant increase
L1.4	-2.52	.012	.65	Significant increase
L2.1	-2.45	.014	.63	Significant increase
L2.5	-2.67	.008	.69	Significant increase
L2.6	-1.89	.059	.49	Not statistically significant
L2.7	-1.34	.180	.35	Not statistically significant
L3.1	-2.13	.033	.55	Significant increase
L3.3	-2.21	.027	.57	Significant increase

Note. Two-sided Wilcoxon signed-rank tests used the Pratt treatment of ties. Effect size was calculated as r = Z/sqrt(N), with N = 15 paired participants. P-values are unadjusted and the analyses are exploratory. The merged L3.6/L3.8 indicator had no direct Round 1 equivalent and was excluded.

4.3 Final Framework Following Delphi Review

4.3.1 Framework Composition

The framework following Delphi review contains 27 indicators. Twenty-six indicators met all prespecified consensus criteria and constitute the consensus-supported core. L2.7, Classroom Dynamics with AI, did not meet the criteria and is retained separately as a provisional research indicator on the basis of the qualitative feedback. It should not be interpreted as a consensus-validated competency or included in a formal total score until further empirical evaluation has been completed.

The consensus-supported core comprises six Foundation indicators, 10 Practice indicators, nine Innovation indicators, and one cross-cutting Comprehensive Ethics indicator. The Practice layer additionally contains L2.7 as the single provisional indicator, resulting in 11 displayed Practice indicators and 27 indicators overall. Thus, 26 of the 27 displayed indicators (96.3%) met the prespecified criteria, while one indicator (3.7%) remained provisional. Figure 2 presents the complete developmental framework.

4.3.2 Final Indicators by Layer

Table 9 presents the final wording and status of each indicator. L3.1 is explicitly role-specific because programme-level curriculum redesign may not fall within the responsibilities of every teacher. L2.7 is visually and analytically separated from the consensus-supported core.

Table 9

AI Literacy Competency Framework Post-Delphi Review

Layer	ID	Final indicator wording
Foundation	L1.1	Understands the possible effects of AI on teaching and learning at the level of higher education.
Foundation	L1.2	Shows an openness to the integration of AI tools in one's teaching.
Foundation	L1.3	Has some understanding of AI and related concepts, such as machine learning, natural language processing, and generative AI.
Foundation	L1.4	Is able to use a range of AI teaching tools, such as supportive teaching AI, teaching content generation AI, and teaching chatbots.
Foundation	L1.5	Has a basic level of data literacy and understands data protection and privacy when using AI tools.
Foundation	L1.7	Knows the relevant Chinese laws, policies, and AI in educational regulations of their institution.
Practice	L2.1	Preparation of learning aims and lesson plans incorporates AI tools and strategies for the enhancement of learning.
Practice	L2.2	Creates or curates AI-embedded teaching resources like adaptive learning materials or interactive simulations.
Practice	L2.3	Utilizes AI interactive teaching tools like real-time Q & A or AI-supported collaborative problem-solving.
Practice	L2.4	Applies AI for the creation and administration of both formative and summative teaching assessments, such as automated/AI graded teaching assessments.
Practice	L2.5	AI is used to provide learners with feedback on their learning and level of achievement that is both individualized and timely.
Practice	L2.6	Relevant AI tools are employed for the appropriate academic discipline.
Practice	L2.7	Understands how to balance classroom dynamics when students use AI. This includes establishing classroom guidelines and mediation of AI-related issues. [Provisional: did not meet consensus criteria]
Practice	L2.8	Makes a judgment on the quality and appropriateness of AI-generated content and the accuracy of such content to inform their practice.
Practice	L2.9	Incorporates AI tools within the relevant systems of the institution including learning management systems.
Practice	N1	Develops assignments and assessments that are inclusive of the use of AI in a transparent and policy compliant way to measure real student learning.
Practice	N2	Uses AI tools to develop and deliver adjustments to teaching that include the use of captions, text to speech, and other alternative formats, to support learners with disabilities.
Innovation	L3.1	Redesigns or reconfigures curriculums or programmes to reflect a meaningful embedding of AI along the learning journey. [Role-specific: primarily applicable to programme directors and equivalent roles]
Innovation	L3.2	Designs and performs cross-disciplinary AI-integrated projects with students of varying academic disciplines.
Innovation	L3.3	Uses AI to enhance the support of the research process including, but not limited to, the searching of literature, the analyzing of research data, and the writing of research, with the appropriate checks and balances for research integrity.
Innovation	L3.4	Uses AI data and evidence to analyze and enhance the quality of teaching.

Note. The framework contains 26 consensus-supported indicators and one provisional indicator (L2.7). N5 operates across all three developmental layers. Indicator wording should remain aligned with the final Delphi audit trail and supplementary records.

Layer	ID	Final indicator wording
Innovation	L3.5	Develops policies pertaining to AI in teaching, learning, and assessment within their department or across the institution.
Innovation	L3.6	Guides or mentors colleagues in AI Literacy, and promotes 'responsible' innovation. [Merged from L3.6 and L3.8]
Innovation	L3.7	Develops ethical frameworks for the teaching context to embed the use of AI with respect to fairness and bias.
Innovation	L3.9	Shares their evidence-based work on the use of AI in the higher education sector through various outlets including writing and speaking.
Innovation	N4	Uses AI to execute teaching material analysis, record review, or student feedback for systematic self-reflection and improvement.
Cross-cutting ethics	N5	Demonstrates ethical practices for teaching with AI, including fairness and equity, transparency, accountability, privacy and data protection, adherence to applicable laws, and respect for academic integrity in consideration of the rights and freedoms of students. [Merged from L1.6 and L1.9]

Note. The framework contains 26 consensus-supported indicators and one provisional indicator (L2.7). N5 operates across all three developmental layers. Indicator wording should remain aligned with the final Delphi audit trail and supplementary records.

4.3.3 Summary of Consensus Status

Delphi identified 26 consensus-supported indicators and 1 consensus-supported provisional indicator. The 26 consensus indicators spanned Foundation (6), Practice (10), Innovation (9), and Comprehensive Ethics (1). L2.7 was kept in the Practice layer to indicate an area which is still developing and needs further exploration. This differentiation is kept in the framework figure, indicator table, Discussion, and Conclusion.

5. Discussion

5.1 Overview of Key Findings

The analysis generated three principal findings. First, the comparative analysis of four Chinese frameworks and two theoretical comparator models revealed substantial heterogeneity and three recurring tensions: standardisation versus contextualisation, static competency lists versus developmental pathways, and instrumental versus agentic conceptions of AI. Second, the Delphi process produced a validated core of 26 indicators organised across three developmental layers—Foundation, Practice, and Innovation—with a Comprehensive Ethics dimension operating across all three layers. During the Delphi rounds, four candidate indicators were considered, three were incorporated, and three structural revisions were implemented. These changes should be reported only if they are fully documented in the Delphi audit trail and supplementary records. Third, L2.7, *Classroom Dynamics with AI*, did not meet the prespecified consensus criteria and showed no statistically significant between-round change (p = .180). It was therefore retained outside the validated core as a provisional research indicator rather than a confirmed competency requirement.

5.2 Dialogue with International Frameworks

The framework shares important features with UNESCO's AI Competency Framework for Teachers, DigCompEdu, and AI-TPACK, while addressing a different contextual and developmental purpose. Table 10 compares their established structures with the present framework.

Table 10

Comparison of the Present Framework With International Frameworks

Framework	Established structure	Relationship to the present framework	Distinctive contribution of the present framework
UNESCO AI Competency Framework for Teachers (2024)	Three progression levels (Acquire, Deepen, Create) for five dimensions and fifteen competencies.	Both frameworks conceptualise teacher AI competence developmentally and foreground human agency and ethical responsibility	The present framework treats ethics as a cross-cutting dimension and operationalises competencies for the Chinese higher education context
DigCompEdu (Redecker, 2017)	Twenty-two competencies across six areas and six proficiency levels	Both frameworks support progressive professional learning, reflection, assessment, and responsible technology use	This framework addresses only AI knowledge, pedagogical practices, ethics, and leadership.
AI-TPACK	While combining knowledge of technology, pedagogy, and content for teaching with AI; Celik (2023) adds ethical knowledge.	Both approaches emphasise the knowledge and judgement required for pedagogically responsible AI integration	The present framework adds an explicit developmental organisation and greater attention to institutional and policy context

5.3 Addressing the Three Core Tensions

The framework responds to the three tensions identified through the comparative analysis. First, it balances standardisation and contextualisation by providing a common developmental structure while allowing discipline- and institution-specific interpretation of individual competencies. Second, it moves beyond a static competency list by organising the indicators across Foundation, Practice, and Innovation layers. These layers should nevertheless be treated as developmental reference points rather than a rigid or universally linear sequence. Third, the proposed Tool → Medium → Co-agent spectrum provides an interpretive account of how teachers' relationships with AI may change as their competence develops. AI supports limited tasks in the Tool stage, mediates pedagogical and cognitive tasks in the Medium stage, and engages in collaborative tasks such as planning, teaching, and reflection in the Co-agent stage where tasks are still controlled by humans. The absence of longitudinal studies regarding the hypothetical Co- Agent stage of AI means that it should be regarded as such and not a sequence of proven stages.

5.4 Theoretical Contributions

This study makes three theoretical contributions to research on higher education teachers' AI literacy.

This framework integrates developmental progression, cross-cutting ethical responsibility, and contextual adaptation in new ways to AI-TPACK. Celik (2023) brought ethical knowledge to Intelligent-TPACK. As such, this framework does not aim to be the first to bring ethics to AI-TPACK. The framework, however, situates ethics as a responsibility invoked at all levels of development. This framework also situates AI literacy as an evolving professional capability rather than a set of knowledge domains (Celik, 2023; Ning et al., 2024).

This framework also applies to the AI-related aspects of professional learning the teacher development models of Fuller (1969) and Huberman (1989). Development, in this case, breaks down to three layers: Foundation, which consists of awareness and basic knowledge; Practice, which is the integration of teaching and the application in the classroom; and Innovation, which is the redesign of curriculum and the engagement in leadership and research, as well the influence on the organization. In this case, the framework is more indicative than prescriptive in that not all teachers develop in the same, linear manner.

Third, the study proposes a Tool → Medium → Co-agent spectrum for conceptualising AI's changing role in education. This spectrum moves beyond a simple distinction between AI as beneficial or threatening and instead considers how human–AI interaction may become more complex as teachers' competencies develop. Because neither UNESCO (2024) nor the

identifiable Eyal (2025) study establishes this exact three-part spectrum, it should be clearly identified as a theoretical proposition developed through the present comparative synthesis.

5.5 Practical Implications

For policymakers, the framework provides an evidence-informed basis for developing future AI-specific competency standards. However, it is essential that the framework is piloted and evaluated before being considered as a basis for the mandatory adoption, certification, or performance assessment. University leaders may use the three descriptive layers to construct differentiated professional development pathways. Teacher educators may use the evidence-based indicators to design the content and outcome of professional development workshops and incorporate ethics in all forms of professional development. The indicators may support the future design of a diagnostic tool, but should not be considered a validated measurement scale. The framework may be used for professional self-appraisal and self-development goal-setting by teachers, but not for outcome-based evaluation.

5.6 The Provisional Indicator: L2.7

L2.7, *Classroom Dynamics with AI*, failed to satisfy the required consensus criteria. Its absence of consensus may indicate if the AI-related classroom dynamics would fall under a separate area of Artificial Intelligence classroom literacy or if they would be considered as part of the general teaching and classroom management skills. The non-significant result ($p = .180$) between the rounds shows that no observable change occurred between rounds; however, this does not imply that it is stable, valid, or that a consensus was reached.

Accordingly, L2.7 should be presented separately from the 26-indicator validated core and clearly labelled as a **provisional research indicator**. It should not contribute to a total framework score or be used as a formal competency requirement until further conceptual clarification and empirical validation have been completed. If qualitative panel comments documented its perceived emerging importance, those comments may be used to explain its provisional retention. Otherwise, the manuscript should state only that the authors retained it for further investigation.

5.7 Limitations

This study has four principal limitations. First, the document analysis included four Chinese frameworks and two theoretical comparator models. Although these sources captured prominent approaches, unpublished or internally developed institutional frameworks may not have been identified. Moreover, the inclusion of author-synthesised comparator models introduces an interpretive element that should be recognised. Second, the Delphi panel was small ($n = 15$), and there was a bias in the representation. The panel included many researchers from a handful of research-intensive universities and specialists in educational technology. This may have limited the representation of diverse perspectives from different disciplines and regions, and from different types of educational and research institutions.. Third, expert consensus only offers initial support for the evidence of relevance and coherence. This evidence does not substantiate other types of validity (e.g., construct, predictive, or criterion). The framework has not been evaluated with adequate sample sizes of teachers, classroom observations or teaching performance, and student outcomes. Fourth, the framework was designed for the Chinese higher education system, and its usability in other largely unresearched national and regional systems and institutional contexts is unknown. Therefore, before making cross-national comparisons, cross-cultural research, including adaptation of the framework and measurement invariance is required.

6. Conclusion and Future Directions

6.1 Summary of the Study

This study addressed the absence of a unified and Delphi-validated competency framework for AI literacy among higher education teachers in China. Using a sequential mixed-methods design—comprising systematic comparative document analysis followed by a two-round Delphi study—we developed a framework containing 27 indicators organised across three developmental layers: Foundation, Practice, and Innovation. A Comprehensive Ethics dimension operates across all three

layers. Twenty-six of the 27 indicators (96.3%) met the prespecified consensus criteria. The remaining indicator, L2.7, Classroom Dynamics with AI, was retained provisionally because experts considered it an emerging competence requiring further empirical investigation. The resulting framework provides an evidence-informed basis for institutional piloting, professional-development planning, policy discussion, and subsequent instrument development.

6.2 Contributions to Theory

This study makes three principal theoretical contributions. First, it integrates AI-TPACK with developmental, ethical, and China-specific contextual dimensions, thereby extending existing accounts of the knowledge required for the pedagogically and ethically responsible use of AI (Celik, 2023; Ning et al., 2024). Second, it contextualises teacher-development stage theory for the AI era by conceptualising AI literacy as a progressive professional capacity rather than a fixed or purely technical competence (Fuller, 1969; Huberman, 1989). Third, it proposes a developmental spectrum for conceptualising AI's role in education, progressing from AI as a Tool to AI as a Medium and, ultimately, AI as a Co-agent. This spectrum illustrates how teachers' relationships with AI may develop from functional use to pedagogical integration and critically governed collaboration.

6.3 Contributions to Practice

The framework offers four practical contributions. First, it provides an evidence-informed foundation for the future development of national or institutional AI competency standards for higher education teachers. Second, its developmental structure can inform tiered professional-development pathways aligned with teachers' differing levels of experience and responsibility. Third, its indicators provide an initial content domain for developing and psychometrically validating diagnostic instruments. Finally, the framework can support structured teacher self-reflection and professional goal-setting. However, it should not yet be used for high-stakes assessment, certification, or performance evaluation until its reliability, validity, and fairness have been established empirically.

6.4 Future Research Directions

Several avenues exist for expanding upon this work. First, while the Delphi method provides an expert consensus on the relevance and organization of the proposed competencies (Hasson et al., 2000; Keeney et al., 2011), it does not address the psychometric validity of the measurement tool. Future studies should therefore implement the framework in the form of a self-assessment or multi-method measurement tool and evaluate its factorial structure, reliability, and validity, both convergent and discriminant, by means of an exploratory and confirmatory factor analysis in research with different samples. Second, a longitudinal and multi-method research design should be used to evaluate the framework's predictive validity in relation to teachers' AI literacy and their subsequent classroom activities, teaching artifacts, students' feedback and educational outcomes. This type of evidence would show the competencies put forth would influence teaching practices rather than the self-perceived competencies of the teachers. Third, the framework should be validated cross-nationally to explore if the framework is conceptually and psychometrically valid in other countries with different education systems and varying policies. This may include a comparative document analysis (Bowen, 2009), adaptation to the local culture, expert reviews, cognitive interviews, and a multi-group confirmatory factor analysis to establish measurement invariance. Lastly, the framework should be evaluated by means of research studies that implement an adapted form of professional development. Best practice would see cluster randomized controlled trials implemented with rigorous design that include comparison groups, randomized designs, and pre, post, and follow-up assessments. These studies would facilitate the movement of the framework into an empirically based tool for research, professional development, and strategic planning for educational institutions.

6.5 Concluding Remarks

Generative AI is reshaping higher education and teachers must construct the pedagogical practices to respond to AI challenges with critical and ethical thinking. This study presents a Delphi-validated pedagogical practices framework. The roadmap layers from basic understanding and application to advanced ethical leadership and innovative practices. The framework should not be viewed as a final goal. The framework is flexible and should be actively modified and improved.

Policymakers and pedagogical leaders should consider frameworks for the continuous growth and development of teachers. The purpose is to provide teachers with the skills and understanding to use AI responsibly and ethically in the classroom to support the learning of students and the development of human agency along the lines of the recent AI and education policies. (UNESCO, 2023, 2024).

Declarations

Author Contribution

Author ContributionsMuhammad Arif: Conceptualisation, methodology, formal analysis, investigation, data curation, writing - original draft, and writing - review and editing.Miracle Uzochukwu Okafor: Methodology, formal analysis, investigation, data curation, and writing - original draft.Mirajur Rhaman Shaoan: Writing - original draft and writing - review and editing.Wang Yong: Conceptualisation, methodology, formal analysis, investigation, data curation, writing - original draft, and writing - review and editing.Bushra Jamil: Writing - original draft and writing - review and editing.

References

1. Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7
2. Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027
3. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
4. 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*. https://doi.org/10.1016/j.chb.2022.107468. Article 107468.
5. Communist Party of China Central Committee, & State Council. (2023). *Plan for the construction of a digital China*. Beijing.
6. Communist Party of China Central Committee, & State Council (2025). *Education powerhouse construction plan (2024–2035)*. Beijing, China. https://english.www.gov.cn/policies/latestreleases/202501/20/content_WS678d85c6c6d0868f4e8eef83.html
7. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*. https://doi.org/10.1186/s41239-023-00392-8. Article 22.
8. Cyberspace Administration of China, National Development and Reform Commission Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, & National Radio and Television Administration. (2023). *Interim measures for the management of generative artificial intelligence services*. Beijing, China.
9. DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.2307/2095101
10. Engestrom, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit.
11. Engestrom, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133–156. https://doi.org/10.1080/13639080020028747
12. Eyal, L. (2025). Rethinking artificial-intelligence literacy through the lens of teacher educators: The adaptive AI model. *Computers and Education Open*, *9*. https://doi.org/10.1016/j.caeo.2025.100291. Article 100291.

13. Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, *6*(2), 207–226. https://doi.org/10.3102/00028312006002207
14. Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x
15. Huberman, M. (1989). The professional life cycle of teachers. *Teachers College Record*, *91*(1), 31–57.
16. Keeney, S., Hasson, F., & McKenna, H. (2011). *The Delphi technique in nursing and health research*. Wiley-Blackwell.
17. Khine, M. S. (Ed.). (2023). Methodology for multilevel modeling in educational research: Concepts and applications. Springer. 10.1007/978-981-16-9144-7
18. Lindell, J., & Stohr, J. (2025). Navigating generative AI in higher education: Six near-future scenarios. *Learning, Media and Technology*. Advance online publication. https://doi.org/10.1080/17439884.2025.2562405
19. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
20. Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340–363. https://doi.org/10.1086/226550
21. Ministry of Education of the People's Republic of China. (2019). *Education modernization 2035*. Beijing.
22. Ministry of Education of the People's Republic of China. (2022). *Teacher digital literacy standard* (pp. 0646–2022). JY/T.
23. Ministry of Education of the People's Republic of China. (2026). *AI + Education action plan*. Beijing. https://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html
24. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
25. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041
26. Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3). Article 978. https://doi.org/10.3390/su16030978
27. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*, n71. https://doi.org/10.1136/bmj.n71
28. Redecker, C. (2017). In Y. Punie (Ed.), *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union. https://doi.org/10.2760/159770
29. Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). Russell Sage Foundation.
30. Salomon, G. (Ed.). (1993). *Distributed cognitions: Psychological and educational considerations*. Cambridge University Press.
31. Sichuan University (2026). Sichuan University teacher AI capacity building action – AI literacy education training full coverage work plan [University work plan]. https://jwc.scu.edu.cn/info/1070/10530.htm
32. State Council of the People's Republic of China. (2017). *Next generation artificial intelligence development plan*. Beijing.
33. Tongji University (2025). *Tongji University teacher AI literacy standard* [Institutional standard]. https://news.tongji.edu.cn/info/1006/92843.htm
34. UNESCO (2019). *Beijing consensus on artificial intelligence and education*. https://www.unesco.org/en/articles/first-ever-consensus-artificial-intelligence-and-education-published-unesco
35. UNESCO (2023). *Guidance for generative AI in education and research*. https://doi.org/10.54675/EWZM9535
36. UNESCO (2024). *AI competency framework for teachers*. https://doi.org/10.54675/ZJTE2084

37. U.S. Department of Education, Office of Educational Technology (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf
38. Zhejiang Provincial Department of Education (2024). *Zhejiang provincial AI literacy framework for higher education teachers* [Institutional handbook] http://jyt.zj.gov.cn/art/2025/9/12/art_1603643_58944381.html

Figures

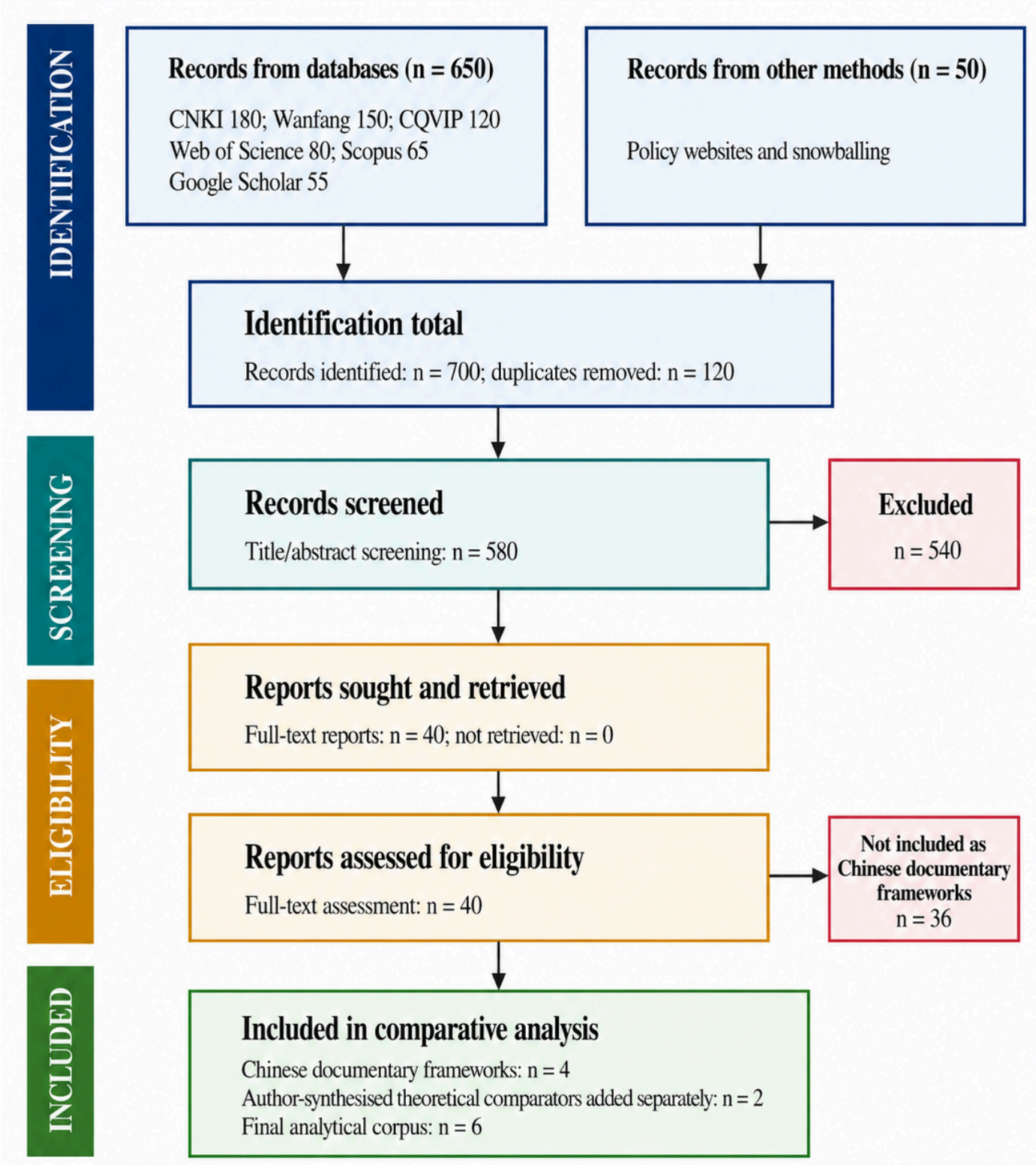


Figure 1

PRISMA-guided identification of four documentary frameworks and the separate addition of two author-synthesised theoretical comparators.

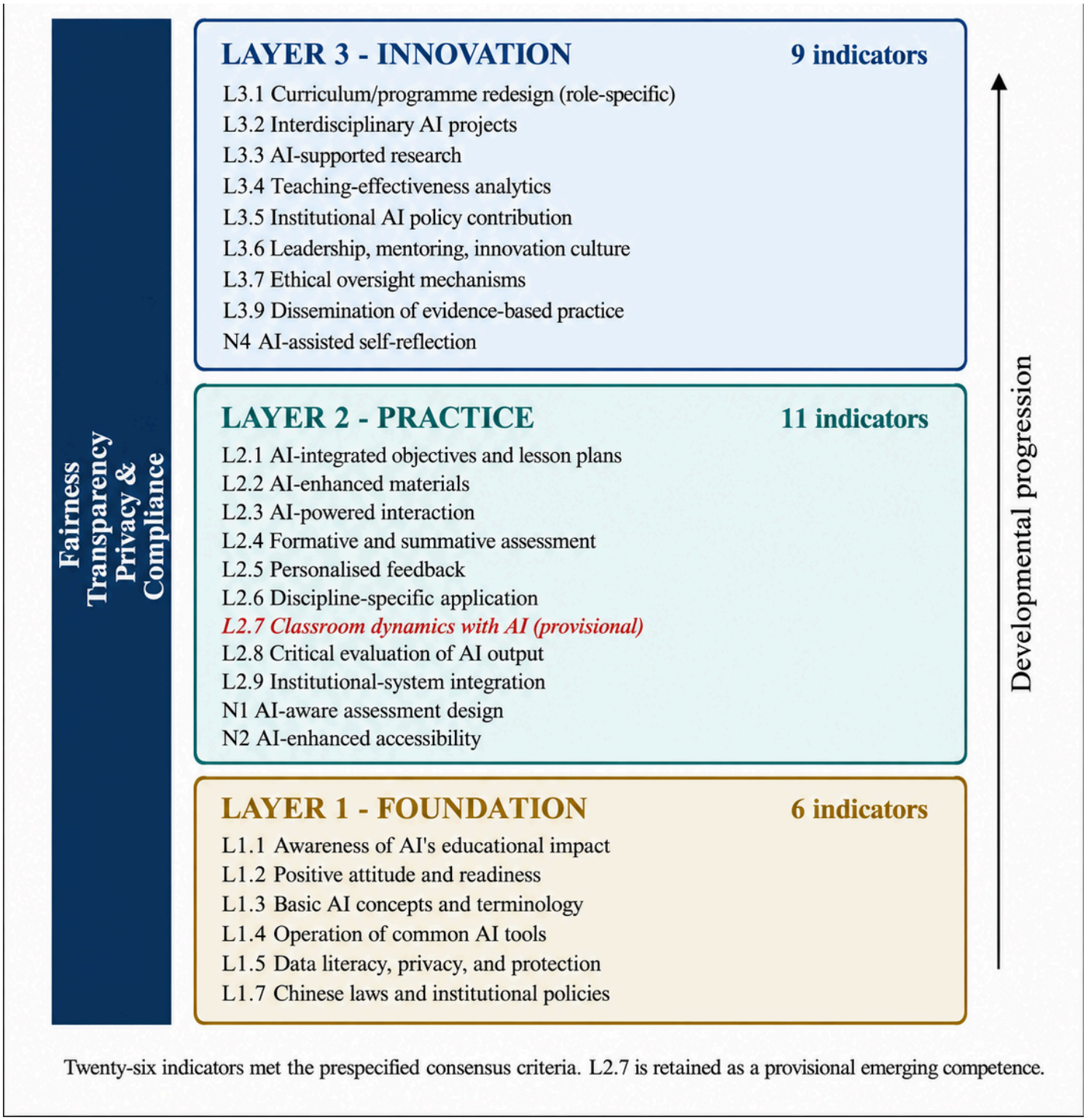


Figure 2

Delphi-validated AI literacy competency framework comprising 26 consensus-supported indicators and one provisional indicator.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialsIJETHEAnonymous.docx