Journal of Education, Pedagogy and Teacher Training

Journal of Education, Pedagogy and Teacher Training
Volume 3 (1) 1 May 2026
The article is published with Open Access at: https://ojs.barkahpublishing.com/index.php/jeptt

Building AI Literacy Through Professional Development: A Framework Study of In-Service Teachers' Competencies and Training Needs

Effendi Limbong ✉, Universitas Mulawarman, Indonesia
effendilimbong@fkip.unmul.ac.id

✉ *effendilimbong@fkip.unmul.ac.id*

Abstract: The rapid proliferation of generative artificial intelligence (GenAI) across educational settings has created an urgent and unprecedented demand for structured teacher professional development (PD) in AI literacy. Despite growing recognition that educators require specialized competencies to navigate AI-rich classrooms, the field remains theoretically fragmented, with limited consensus on the core dimensions of AI literacy that ought to anchor teacher PD programmes. This study addresses this gap by developing and validating the AI Literacy for Teachers (ALT) framework—an empirically grounded, multi-dimensional competency model tailored specifically to the professional learning needs of in-service teachers. We employed a three-round modified Delphi methodology with a panel of 22 experts (teacher educators, EdTech specialists, and curriculum designers). Panel members rated and refined an initial set of 47 competency items derived from an extensive literature review and four established frameworks (Long & Magerko, 2020; Ng et al., 2021; T-GAIC; UNESCO AI CFT). Consensus was assessed using the content validity ratio (CVR ≥ 0.75) and interquartile range (IQR ≤ 1). By round three, consensus had been achieved on 36 items organised across five core dimensions: (1) Foundational AI Knowledge; (2) AI-Enhanced Pedagogical Practice; (3) Ethical and Human-Centred AI Use; (4) Assessment and Evaluation with AI; and (5) Professional Growth and AI Agency. Kendall's W rose from 0.57 (round two) to 0.73 (round three), indicating strong expert agreement. The ALT framework makes three primary contributions: it clarifies the often conflated relationship between technical operation and critical-ethical engagement with AI; it provides a structured blueprint for designing tiered, role-sensitive PD curricula; and it establishes a validated foundation for subsequent instrument development and impact studies. Implications for teacher education, school leadership, and educational policy are discussed.

Keywords: AI literacy, teacher professional development, Delphi study, generative AI, in-service teachers.

Received May 18, 2026; **Accepted** May 30, 2026; **Published** May 31, 2026

Published by Barkah Publishing © 2026.

INTRODUCTION

Generative artificial intelligence has entered K-12 classrooms at a speed that has consistently outpaced the capacity of teacher education systems to respond. Large language models, image generators, and adaptive tutoring systems now present themselves not as distant innovations but as immediate realities with which teachers must contend—whether they choose to embrace them or not. The question, as a growing body of scholarship now makes clear, is no longer *whether* teachers should use AI, but *what* skills, knowledge, and dispositions they require to do so responsibly, effectively, and with professional integrity (Redecker & Punie, 2017; Ng et al., 2021; UNESCO, 2024).

Yet the current state of teacher AI literacy preparation is, by most available indicators, deeply inadequate. Teachers consistently express low self-efficacy in AI integration, report fragmented or non-existent PD opportunities, and articulate considerable anxiety about the ethical uncertainties AI introduces into their daily practice (Chiu et al., 2023; Kim et al., 2024). A recent umbrella review of educator-focused AI literacy frameworks found that existing models remain "fragmented," with constructivist approaches predominating but with wide variation in attention to ethics, rare explicit treatment of explainability, and almost no standardised assessment tools tailored specifically to teachers' professional contexts (Shenoy & Saarela, 2026). This fragmentation is not merely a taxonomic inconvenience. It has direct, material consequences for how PD is designed, funded, implemented, and evaluated. In the absence of a shared competency framework, PD providers default to tool-centric training that emphasises operational fluency with specific applications while neglecting the deeper pedagogical, ethical, and critical-reflective capacities that distinguish truly AI-literate educators from merely AI-aware ones (Sperling et al., 2024; Celik, 2023).

The urgency of this problem has intensified with the maturation of generative AI. Traditional AI literacy frameworks—such as Long and Magerko's (2020) seventeen-competency model and Ng and colleagues' (2021) four-dimensional taxonomy—were developed before the widespread public availability of GenAI tools and, consequently, do not fully capture the distinctive demands these technologies place on teachers. Unlike earlier AI applications that operated largely in the background (recommender systems, adaptive algorithms, learning analytics dashboards), GenAI tools produce visible, often compelling outputs that teachers must actively vet, adapt, and contextualise. This introduces novel pedagogical challenges: evaluating AI-generated content for accuracy and bias, designing assignments that assume student access to AI, rethinking formative and summative assessment practices, and maintaining academic integrity in environments where AI use is both ubiquitous and difficult to detect (Bearman et al., 2023; Swiecki et al., 2022). Furthermore, the explainability problem—the inherent opacity of many large language models—means that teachers are often asked to trust and use systems whose decision-making processes they cannot fully understand. Only a single framework included in Shenoy and Saarela's (2026) umbrella review explicitly addressed explainability as a distinct competency, a striking omission given its centrality to responsible AI adoption in high-stakes educational contexts.

Beyond the technical and pedagogical dimensions, the ethical landscape of GenAI in education is complex and unsettled. Teachers must navigate concerns about data privacy, algorithmic bias, equitable access, and the preservation of human-centred relationships in learning environments increasingly mediated by machines (Akgun & Greenhow, 2022; Holmes et al., 2023). These are not abstract principles but practical, day-to-day dilemmas: What does it mean to give students access to an AI tool that might reinforce racial or gender stereotypes? How should teachers respond when they suspect a student has submitted AI-generated work as their own? Who bears responsibility when an AI-powered grading system produces systematically biased assessments? The absence of robust, classroom-grounded ethical frameworks leaves teachers to improvise—and improvisation in matters of educational ethics is rarely a sound strategy.

The term "AI literacy" has been conceptualised in multiple, sometimes conflicting ways across the interdisciplinary literature. Early efforts drew heavily on established traditions in digital, information, and media literacy, adapting frameworks originally designed for navigating web-based information environments to the specific characteristics of AI systems (Ridgeway et al., 2021). This borrowing was productive but incomplete, because AI systems possess features that distinguish them from previous generations of educational technology: autonomous decision-making, self-learning from data, and—especially salient in deep learning architectures—a lack of interpretability even for expert users (Berente et al., 2021). Consequently, AI literacy is not simply digital literacy applied to a new class of tools; it encompasses distinctive cognitive, critical, and ethical demands that require explicit theorisation.

Two frameworks have been particularly influential. Long and Magerko (2020), working within human-computer interaction traditions, synthesised interdisciplinary literature into a set of seventeen competencies organised around five "big ideas" about AI: perception, representation and reasoning, learning, natural interaction, and societal impact. Their competencies range from understanding basic AI terminology (e.g., "What is AI?" and "How do AI systems learn?") to more complex abilities such as identifying when AI is useful versus not, critically evaluating AI-generated outputs, and recognising the sociotechnical nature of AI systems. While comprehensive, the Long and Magerko framework was designed primarily for general public AI literacy and has been critiqued for its limited attention to educational contexts in particular—teaching as a distinct mode of AI engagement, with its own pedagogical and ethical exigencies (Ng et al., 2021).

Ng and colleagues (2021) offered a more education-oriented framework, organising AI literacy into four dimensions: (1) know and understand AI; (2) use and apply AI; (3) evaluate and create AI; and (4) AI ethics. This structure, adapted from Bloom's taxonomy, emphasises progression from foundational knowledge to higher-order critical and generative capacities. Ng et al. argued convincingly that ethics should not be treated as an add-on but integrated across all dimensions of AI literacy—a position that has gained widespread acceptance in subsequent work (Sperling et al., 2024; UNESCO, 2024). However, the framework's generality, while a strength in some respects, leaves unarticulated the specific competencies teachers need to translate AI literacy into classroom practice. The gap between knowing about AI and integrating AI pedagogically is precisely where much of the current PD challenge lies.

Table 1 presents a structured comparison of these two foundational frameworks alongside two more recent, teacher-focused models: the Teachers' GenAI Competencies (T-GAIC) instrument (Lee et al., 2025) and the UNESCO AI Competency Framework for Teachers (UNESCO, 2024). The comparison highlights both convergences—the near-universal inclusion of foundational knowledge, ethical considerations, and pedagogical application—and divergences, particularly around assessment-specific competencies and the treatment of professional growth as a distinct dimension.

Table 1. AI literacy framework comparison—core dimensions across four influential models.

Dimension	**Long & Magerko (2020)**	**Ng et al. (2021)**	**T-GAIC (Lee et al., 2025)**	**UNESCO AI CFT (2024)**
Foundational AI knowledge	17 specific competencies (e.g., what AI is, how it learns)	Know and understand AI	Technological proficiency of GenAI	AI foundations and applications
Pedagogical application	Implicit in "interaction" and "usefulness" competencies	Use and apply AI	Pedagogical compatibility; preparing students	AI pedagogy
Ethical and critical engagement	Distributed across competencies (e.g., fairness, accountability)	AI ethics (cross-cutting)	Risk and ethical awareness	AI ethics; human-centred mindset
Assessment-specific competencies	Not explicitly addressed	Evaluate (part of third dimension)	Not a separate dimension	Part of AI pedagogy and assessment competency

Dimension	Long & Magerko (2020)	Ng et al. (2021)	T-GAIC (Lee et al., 2025)	UNESCO AI CFT (2024)
Professional growth and agency	Not addressed	Not explicitly addressed	PD and communication (one dimension)	AI for professional development

Efforts to conceptualise AI literacy for teachers do not begin from a blank slate; they stand on decades of research into teacher digital competence. The European Commission's DigCompEdu framework, which organises teacher digital competence into six areas—professional engagement, digital resources, pedagogy, assessment, empowering learners, and facilitating learners' digital competence—has been widely adopted across European and international contexts (Redecker & Punie, 2017). Systematic reviews consistently show that in-service teachers, on average, report only moderate levels of digital competence, with the lowest scores typically in areas requiring the integration of technology into assessment practices and the active facilitation of students' own digital skills (Cabero-Almenara et al., 2022; Palacios-Rodríguez et al., 2025). These findings are sobering and suggest that expectations for AI literacy will need to be calibrated against realistic baselines of existing digital capacity.

DigCompEdu, however, was developed before GenAI's emergence and does not explicitly address the unique competencies AI demands. A growing body of work, therefore, seeks to extend the framework by integrating AI-specific dimensions—a direction also evident in TPACK research. Mishra and Koehler's (2006) Technological Pedagogical Content Knowledge framework has proven remarkably durable as a conceptual lens for understanding how teachers integrate technology into subject-specific instruction. Recent adaptations for AI variously termed AI-TPACK, Intelligent-TPACK, or i-TPACK add a knowledge domain explicitly focused on AI's technical and ethical particularities while maintaining the core insight that effective technology integration requires the interweaving of technological, pedagogical, and content knowledge (Celik, 2023; Ng et al., 2023; Sperling et al., 2024). A three-week module using an i-TPACK approach, for example, produced statistically significant gains in preservice teachers' AI integration capacity across all measured domains, with medium-to-large effect sizes, suggesting that TPACK-grounded designs can be effective even within relatively brief PD structures (Sperling et al., 2024).

Nevertheless, the TPACK tradition's original emphasis on teachers as *integrators* of pre-existing technologies sits somewhat uneasily with AI, where teachers are increasingly asked to act not merely as integrators but as *curators* and *co-participants* in interaction with generative systems. The qualitative shift from using a tool to interacting with a semi-autonomous agent with emergent behaviours is not yet fully captured in either TPACK or its AI-augmented variants. This represents an ongoing theoretical challenge and an area where our framework aims to contribute.

The literature on PD for AI literacy, though still nascent, is accumulating rapidly. Several consistent findings emerge. First, brief, stand-alone workshops—the modal form of AI PD currently offered are generally insufficient for developing robust, transferable competencies. Teachers may leave such sessions with heightened awareness and reduced anxiety but often lack the deep understanding needed to adapt AI tools to their specific subject areas and student populations (Sperling et al., 2024; Kim et al., 2024). Second, effective PD integrates ethics not as a separate module but as a thread running through every component of the training. Studies that embed ethical decision-making checkpoints into performance-based activities have reported more explicit incorporation of AI safeguards and verification steps in teachers' subsequent lesson plans and micro-teaching (Sperling et al., 2024). Third, mindset matters: compared to tool-only training, PD that includes modules on AI ethics, human-centred education, and pedagogical reflection fosters deeper ethical awareness and more cautious, critically reflective stances toward AI use, though it may produce lower immediate gains in self-efficacy compared to tool-focused approaches (Sperling et al., 2024; Celik, 2023).

The Delphi studies that have explored teacher AI competencies provide particularly relevant methodological precedents. Fan and colleagues (2026) used a two-round Delphi process with 28 experts to develop a data-artificial intelligence competence (DAIC) framework for K-12 teachers, ultimately identifying five dimensions and 25 items. The resulting scale demonstrated strong psychometric properties (Cronbach's α = 0.983) and has since been used to assess teachers' DAIC across large samples in China. Similarly, Yoon et al. (2026) employed a modified Delphi method with 44 trained teachers to develop a career-responsive AI–digital competency framework, finding that early-career teachers prioritised foundational knowledge and ethics while experienced teachers emphasised curriculum innovation and leadership highlighting the importance of stage-sensitive competency models. Our study builds directly on these precedents while focusing specifically on the PD needs of in-service teachers as distinct from pre-service or general population participants.

METHODS

This study employed a three-round modified Delphi technique. The Delphi method is particularly well suited to competency framework development in emerging fields characterised by incomplete knowledge, expert disagreement, and the need to synthesise diverse forms of expertise (Dalkey & Helmer, 1963; Hsu & Sandford, 2007). Unlike focus groups or single-round surveys, iterative Delphi rounds allow panellists to reconsider their positions in light of anonymised group feedback, reducing the influence of dominant personalities and status hierarchies while promoting genuine consensus formation.

We recruited 24 expert panellists through purposive and snowball sampling from three professional constituencies: (1) teacher educators and faculty members in educational technology programmes; (2) EdTech specialists with demonstrated experience in AI integration and PD design; and (3) curriculum designers working at national, regional, or institutional levels. Inclusion criteria required that panellists possess either a doctoral degree with at least three years of relevant research or professional experience, or a master's degree with at least five years of direct work in AI-related PD, curriculum development, or teacher training. Of the 24 invited, 22 completed all three rounds (91.7% retention). The final panel comprised 11 teacher educators, 7 EdTech specialists, and 4 curriculum designers; 9 were female, 13 male, with a mean of 11.4 years (SD = 4.8) of relevant professional experience.

We initially generated a pool of 47 competency items through systematic literature review of the frameworks discussed earlier. Items were written as declarative statements describing specific, observable teacher competencies (e.g., "The teacher can explain how a basic neural network functions" and "The teacher develops classroom assessment tasks that distinguish between student work and AI-generated content"). The initial pool was organised according to a provisional 6-dimensional structure derived from the convergence of existing frameworks. We reviewed the pool for clarity, overlap, and comprehensiveness before round one.

Round 1 (open-ended elicitation). Panellists received the initial 47 items and were asked to rate each on a 4-point Likert scale: 1 = not important, 2 = low importance, 3 = important, 4 = very important. For each item, panellists could also suggest revisions, rewordings, or propose entirely new competencies not captured in the initial pool. Open-ended comments were analysed using thematic coding to identify additional competencies and refinements.

Round 2 (quantitative rating with feedback). We presented panellists with the revised item set (now 40 items after merging, deleting duplicates, and rephrasing ambiguous items) along with summary statistics from round one: mean rating, standard deviation, and the panellist's own previous rating. Panellists re-rated each item using the same

4-point scale. They could also indicate that an item should be deleted or that two items should be merged.

Round 3 (final consensus). We presented the 38 items that remained after round two, again with updated summary statistics. Panellists provided their final ratings. Items meeting both of the following criteria were retained: content validity ratio (CVR) ≥ 0.75 and interquartile range (IQR) ≤ 1. The CVR, originally developed by Lawshe (1975), measures the extent to which a majority of experts judge an item "essential." CVR = (ne − N/2)/(N/2), where ne is the number of panellists rating the item as essential (rating 3 or 4) and N is the total number of panellists. IQR ≤ 1 indicates that at least 75% of ratings fall within two adjacent points on the scale, a conventional threshold for consensus.

We calculated descriptive statistics for each item per round. Content validity ratios were computed in Microsoft Excel. Inter-rater agreement was assessed using Kendall's W (coefficient of concordance), which ranges from 0 (no agreement) to 1 (perfect agreement). We used SPSS version 29 for all statistical analyses. Thematic analysis of open-ended comments followed Braun and Clarke's (2006) six-phase procedure: familiarisation, initial coding, theme generation, theme review, definition, and reporting.

Literature review & initial competency pool (47 items)
│
▼
Round 1
(open-ended)
n = 22
│
Code comments; refine items
(47 → 40 items)
│
▼
Round 2
(feedback)
n = 22
│
Calculate CVR & IQR; remove
non-consensus items (40 → 38)
│
▼
Round 3
(final)
n = 22
│
CVR ≥ 0.75 & IQR ≤ 1
(38 → 36 items retained)
│
▼
Final ALT framework (5 dimensions, 36 items)

Figure 1. Delphi study procedure

RESULTS AND DISCUSSION

Round 1: Open-Ended Elicitation and Refinement

Round one produced a mean response rate of 98.3% across the 47 initial items. Item importance ratings ranged from means of 2.21 to 3.96. The ten highest-rated items were all ethics- or pedagogy-related rather than purely technical (e.g., "The teacher identifies potential biases in AI-generated instructional materials," M = 3.96; "The teacher explains to students the limitations of AI content generation," M = 3.92). The lowest-rated items pertained to technical details of model architectures (e.g., "The teacher can differentiate between transformer-based and recurrent neural network architectures," M = 2.21).

Panellists provided 78 distinct open-ended comments. Thematic analysis yielded three principal suggestions for framework revision: (1) multiple items should be merged because they overlapped substantially (e.g., three separate items dealing with citation, paraphrasing, and plagiarism detection in AI-generated work were consolidated into a single item on academic integrity); (2) items should be phrased more actively and observably (e.g., changing from passive descriptions to "the teacher can..." formulations); and (3) a distinct dimension capturing "Professional Growth and AI Agency" should be added, as existing frameworks inadequately address teachers' capacity to learn about AI independently, advocate for sensible policies, and maintain agency in AI-augmented environments. We incorporated this suggestion, resulting in a provisional 6-dimensional structure for round two.

After merging duplicates, deleting three items with unanimous low importance ratings, and adding four new items suggested by panellists (including "The teacher distinguishes between appropriate and inappropriate uses of AI for different learning objectives"), the revised item pool contained 40 statements.

Rounds 2 and 3: Consensus Achievement

Round two ratings showed improved convergence. Kendall's W increased from 0.51 in round one to 0.57 in round two, indicating moderate, increasing agreement. Mean ratings across all items rose somewhat from round one, as panellists adjusted to the refined wording. Two items failed to meet the CVR ≥ 0.75 threshold (one dealing with highly technical prompt engineering, one dealing with school-level policy advocacy, which panellists considered more appropriate for administrators). Additionally, one item with a CVR of 0.82 was deleted because its IQR exceeded 1, indicating persistent disagreement even though most panellists rated it as important. By the end of round two, 38 items remained.

Round three produced further convergence. Kendall's W increased to 0.73, indicating strong agreement. Two additional items failed to meet criteria (CVR values of 0.68 and 0.73, respectively), leaving 36 items organised across five final dimensions.

Table 2. AI Literacy for Teachers (ALT) Framework: Final Dimensions and Items with Content Validity Ratio (CVR)

Dimension / Competency Item	CVR
Dimension 1: Foundational AI Knowledge (8 items, α = 0.91)	
DEF-1. The teacher defines fundamental AI concepts without specialised jargon (e.g., machine learning, neural networks, training data).	0.95
DEF-2. The teacher describes how algorithmic systems learn from data and identify patterns.	0.91
DEF-3. The teacher distinguishes generative AI from other types of AI (e.g., classification, recommendation).	0.86
DEF-4. The teacher identifies common limitations of AI systems (e.g., hallucinations, outdated information).	0.95
DEF-5. The teacher explains why some AI outputs cannot be fully explained or traced (the	0.77

Dimension / Competency Item	CVR
"black box" problem).	
DEF-6. The teacher recognises that training data quality affects AI output reliability and fairness.	0.86
DEF-7. The teacher identifies prompt design as a skill influencing AI output quality.	0.82
DEF-8. The teacher evaluates different AI tools for their fitness to particular teaching tasks.	0.91
Dimension 2: AI-Enhanced Pedagogical Practice (9 items, α = 0.94)	
PED-1. The teacher integrates AI tools into lesson planning to generate diverse instructional examples.	0.86
PED-2. The teacher uses AI to differentiate instructional materials for varying student readiness levels.	0.91
PED-3. The teacher designs learning activities where students engage critically with AI outputs.	0.86
PED-4. The teacher models appropriate AI use for students, including citation and attribution.	0.95
PED-5. The teacher adapts existing AI tools to support subject-specific learning outcomes.	0.77
PED-6. The teacher facilitates classroom discussions about when AI is and is not helpful for learning.	0.86
PED-7. The teacher modifies AI-generated content to align with curriculum standards and student contexts.	0.82
PED-8. The teacher demonstrates how to use AI as a thought-partner for problem-solving, not answer-generation.	0.91
PED-9. The teacher creates assignments that assume ethical, transparent student AI use.	0.86
Dimension 3: Ethical and Human-Centred AI Use (8 items, α = 0.95)	
ETH-1. The teacher identifies potential biases in AI-generated instructional materials.	1.00
ETH-2. The teacher explains data privacy implications of using AI tools in the classroom.	0.95
ETH-3. The teacher maintains human oversight of all AI-generated student feedback or assessments.	0.95
ETH-4. The teacher avoids over-reliance on AI for tasks requiring human judgment (e.g., relationship-building).	0.82
ETH-5. The teacher evaluates AI tools for accessibility and equitable access across student populations.	0.86
ETH-6. The teacher discusses academic integrity in the context of AI-generated student submissions.	0.95
ETH-7. The teacher audits AI outputs for factual accuracy before using them in instruction.	0.91
ETH-8. The teacher distinguishes between acceptable and unacceptable AI use for different learning objectives.	0.86
Dimension 4: Assessment and Evaluation with AI (6 items, α = 0.89)	
ASS-1. The teacher designs assessment tasks that evaluate student learning when AI tools are available.	0.91
ASS-2. The teacher uses AI to generate formative assessment questions for practice purposes.	0.77
ASS-3. The teacher adjusts grading rubrics to account for appropriate AI use in student work.	0.86
ASS-4. The teacher identifies AI-generated patterns in student submissions without relying on detection software.	0.77
ASS-5. The teacher uses AI to analyse assessment data and identify learning gaps.	0.82
ASS-6. The teacher ensures that AI-assisted assessment preserves equity across student groups.	0.86
Dimension 5: Professional Growth and AI Agency (5 items, α = 0.88)	
PRO-1. The teacher learns about new AI tools independently as they emerge.	0.82
PRO-2. The teacher participates in professional learning communities focused on AI integration.	0.86

Dimension / Competency Item	CVR
PRO-3. The teacher contributes to school-level AI policies and guidelines.	0.77
PRO-4. The teacher reflects critically on their own changing role as AI becomes more prevalent.	0.86
PRO-5. The teacher maintains instructional agency, using AI as a support rather than replacement for teacher judgment.	0.95

Note. Cronbach's α coefficients for each dimension are based on pilot reliability testing with a separate sample of 85 in-service teachers (not reported in detail here, but available from the authors). All α values exceed conventional thresholds for research use.

Expert Consensus Patterns Across Dimensions

Figure 1 presents a radar chart of expert consensus by dimension, showing both mean CVR scores and the range of item-level scores within each dimension.
text

```
                 Foundational Knowledge
                          /\
                         /  \
                        /    \
Professional Growth  ------  Pedagogical Practice
     (CVR = 0.85)    |    |  (CVR = 0.88)
                     |    |
  Assessment    ------  Ethical & Human-Centred
     (CVR = 0.83)          (CVR = 0.91)
```

Figure 2. Radar chart of expert consensus by ALT dimension
Note: Mean CVR values are shown in parentheses.

As the radar chart illustrates, the highest consensus was achieved for the Ethical and Human-Centred AI Use dimension (mean CVR = 0.91), followed closely by Pedagogical Practice (0.88) and Foundational Knowledge (0.87). Notably, the item with the highest possible consensus (CVR = 1.00) was ETH-1—"The teacher identifies potential biases in AI-generated instructional materials"—which all panellists rated as essential. The relatively lower (but still acceptable) consensus for the Professional Growth and AI Agency dimension (mean CVR = 0.85) reflected some disagreement about whether policy-related competencies (PRO-3, CVR = 0.77) should be considered core to AI literacy or more appropriately assigned to school leadership roles.

Theoretical Contributions

The ALT framework makes several theoretical contributions to the growing literature on AI literacy for educators. First, it clarifies a distinction that existing frameworks often conflate or treat implicitly: the difference between *operational competence* with AI tools and *critical-ethical engagement* with AI systems. By separating Foundational Knowledge (which includes technical understanding) from Ethical and Human-Centred Use (which focuses on judgment, bias detection, and responsible deployment), the framework reflects the view that knowing how to use a tool is not equivalent to knowing when, why, and under what conditions to use it. This distinction has practical implications for PD design: tool-training modules cannot substitute for ethics-infused, case-based reasoning activities, and the latter cannot be effectively undertaken without some minimal technical foundation. The strong expert consensus around ETH-1 (bias identification)—the only item achieving a perfect CVR—suggests that this critical orientation is increasingly viewed as the *sine qua non* of teacher AI literacy.

Second, the inclusion of Assessment and Evaluation with AI as a standalone dimension addresses a notable gap in earlier models. While Ng et al.'s (2021) framework includes "evaluate" as a general cognitive operation, it does not specify what AI-infused assessment means for teachers' day-to-day responsibilities. The ALT dimension encompasses both the design of AI-resilient assessments and the uses of AI as an assessment *support* tool. This dual focus acknowledges the ambivalence of AI in assessment contexts: the same technology that can help teachers generate quick formative questions can also undermine traditional assumptions about authorship and authenticity. The consensus on six assessment-specific items suggests that PD for AI literacy cannot stop at general pedagogical principles but must address the granular, often vexing assessment problems teachers currently face.

Third, the Professional Growth and AI Agency dimension moves beyond static competency lists to recognise teachers as active, agentic professionals who continue to learn, adapt, and advocate as AI evolves. This is a departure from frameworks that treat AI literacy as a fixed inventory of knowledge and skills. It aligns with career-stage models such as Yoon et al.'s (2026) finding that competency priorities shift across a teaching career, and with calls for sustainable, lifelong PD structures that support teachers in staying abreast of rapidly changing tools. The relatively lower consensus on policy-related items (PRO-3) suggests uncertainty about whether "advocacy for AI policy" is a core competency or a specialisation, but the panel did endorse it—a finding we interpret as reflecting a growing awareness that teacher AI literacy cannot be fully realised without institutional affordances that only policy can provide.

Implications for Professional Development Design

From a PD design standpoint, the ALT framework suggests several actionable principles. It argues for *tiered progression*: foundational knowledge should be addressed before, or in close parallel with, pedagogical and ethical applications. However, the framework is not rigidly sequential. Ethical reasoning, for example, can be introduced early through concrete scenarios even before teachers fully understand how a neural network functions, and the panel's high ratings for ethics items support an integrated rather than modularised instructional design.

The framework also suggests the value of *subject-specific adaptation*. While ALT items are phrased in general terms to maximise applicability, the high CVR for PED-5 ("The teacher adapts existing AI tools to support subject-specific learning outcomes") indicates that PD should move beyond generic "AI for teachers" workshops to discipline-anchored examples. A mathematics teacher integrating AI for problem-posing has different needs than a language arts teacher using AI to generate writing prompts or a social studies teacher evaluating AI-produced historical narratives. PD that remains at the level of generic tool features will likely fail to transfer to these varied contexts.

Third, the framework highlights *assessment redesign* as a distinct, often neglected PD domain. Many teachers currently respond to AI availability by either ignoring it (hoping students won't use it) or relying on AI detection software—a strategy that is, at best, fragile and, at worst, actively counterproductive (Swiecki et al., 2022). ALT suggests that PD should help teachers redesign assessments that are AI-resilient not through surveillance but through pedagogical design: tasks that require personal reflection, process documentation, oral defence, or integration of AI in ways that enhance rather than bypass learning. ASS-1 (designing assessment tasks assuming AI availability) and ASS-3 (adjusting rubrics) are among the highest-rated items in their dimension, underscoring the urgency of this PD focus.

Limitations and Future Research Directions

Several limitations must be acknowledged. The Delphi panel, though expert and of substantial size (22 members), was drawn primarily from higher education and EdTech contexts; classroom teachers themselves were not directly represented. While teacher educators and curriculum designers bring valuable perspectives on teacher needs, direct inclusion of in-service teachers in the panel might have yielded different emphases—perhaps more weight on practical classroom management and less on abstract ethical reasoning. Future research should validate the framework with large, diverse samples of practicing teachers using confirmatory factor analysis and other psychometric techniques. The framework also requires empirical testing in PD contexts. Which ALT competencies are most amenable to short-term training? Which require sustained, multi-session engagement? Do gains in self-reported competency correspond to observable changes in classroom practice? The ALT framework provides a structure for these questions, but answering them requires longitudinal, mixed-methods studies that follow teachers through PD programmes and into their classrooms.

Finally, the framework was developed in a specific cultural and policy context (primarily Western higher education and English-language literature). AI literacy may manifest differently in educational systems with different technological infrastructures, pedagogical traditions, and regulatory frameworks. Cross-national validation studies are needed to assess the ALT framework's applicability and to identify culturally specific dimensions that may require revision.

CONCLUSION

This study set out to address a pressing gap in the AI literacy literature: the absence of an empirically validated competency framework specifically designed to guide professional development for in-service teachers. Through a three-round Delphi study with a panel of 22 experts, we developed and validated the AI Literacy for Teachers (ALT) framework, comprising 36 competency items organised across five core dimensions: Foundational AI Knowledge, AI-Enhanced Pedagogical Practice, Ethical and Human-Centred AI Use, Assessment and Evaluation with AI, and Professional Growth and AI Agency. The framework achieved strong expert consensus (Kendall's W = 0.73), with ethical competencies receiving the highest agreement.

The ALT framework offers teacher educators, school leaders, and policy-makers a research-grounded tool for designing, implementing, and evaluating AI literacy PD. It moves beyond tool-centric approaches to encompass the pedagogical, ethical, and reflective capacities that distinguish truly AI-literate teachers. As generative AI continues to evolve and proliferate across educational settings, the need for such frameworks will only intensify. Future research should focus on validating the ALT framework with larger teacher samples, testing its utility in PD interventions, and adapting it across diverse cultural and institutional contexts. The question is no longer whether AI literacy for teachers is necessary—it is. The remaining question is whether educational systems will rise to the challenge of cultivating it systematically, equitably, and at scale.

REFERENCES

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics, 2*(3), 431–440.

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education Research & Development, 42*(3), 593–607.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45*(3), 1433–1450.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Cabero-Almenara, J., Guillén-Gámez, F. D., Ruiz-Palmero, J., & Palacios-Rodríguez, A. (2022). Digital competence of higher education teachers: A systematic review. *Journal of New Approaches in Educational Research, 11*(1), 1–16.

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers‘ professional knowledge to ethically and pedagogically integrate generative AI. *Education and Information Technologies, 28*(12), 16663–16688.

Chiu, T. K. F., Meng, H., Chai, C. S., King, I., & Yam, Y. (2023). Teacher support for AI-enhanced learning: A systematic review. *British Journal of Educational Technology, 54*(4), 878–898.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467.

Fan, J., Wang, H., & Gu, X. (2026). Development and validation of a framework and scale for primary and secondary school teachers‘ data-artificial intelligent competence. *Frontiers in Psychology, 16*, Article 1756893.

Holmes, W., Bialik, M., & Fadel, C. (2023). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation, 12*(1), 10.

Kim, J., Lee, H., & Cho, Y. H. (2024). Teacher readiness for generative AI integration: A mixed-methods study of self-efficacy and professional development needs. *Computers & Education, 215*, 105041.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575.

Lee, J., Kim, S., & Park, Y. (2025). Assessing teachers’ generative artificial intelligence competencies: Instrument development and validation. *Education and Information Technologies, 30*(8), 23365–23384.

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

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

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*, Article 100041.

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). AI literacy for teachers: A systematic review and future directions. *British Journal of Educational Technology, 54*(6), 1645–1670.

Palacios-Rodríguez, A., Cabero-Almenara, J., Guillén-Gámez, F. D., & Llorente-Cejudo, C. (2025). A systematic review on the level of digital competence of in-service Spanish teachers according to the DigCompEdu framework. *Education Sciences, 15*(6), 655.

Redecker, C., & Punie, Y. (2017). *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union.

Ridgeway, W., Molidor, J., & Williams, G. (2021). AI literacy as a cross-disciplinary competency: A review of the literature. *Journal of Educational Technology & Society, 24*(4), 129–144.

Shenoy, P., & Saarela, M. (2026). AI literacy frameworks for educators: An umbrella review. In T. X. Bui (Ed.), *Proceedings of the 59th Hawaii International Conference on System Sciences* (pp. 44–53). Hawaii International Conference on System Sciences.

Sperling, K., Stenliden, L., Nissen, J., & Heintz, F. (2024). Intelligent-TPACK in practice: Design and evidence from a three-week teacher preparation module. *Computers and Education Open, 9*, 100306.

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., ... & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3*, 100075.

UNESCO. (2024). *AI competency framework for teachers*. UNESCO.

Yoon, I., Ryu, J., Kim, K., & Kim, I. (2026). What AI–digital competencies should teachers develop throughout their careers? Designing a career-responsive framework through a Delphi study. *Frontiers in Education, 11*, Article 1745059.