Mapping AI Literacy and AI-Related Digital Competencies in U.S. K–12 Teacher Professional Development: A Systematic Review and Bibliometric Analysis

Mariel Evelyng Castellanos Adarme[1], Luis Eduardo Muñoz Guerrero[2], Efrén A. Báez Conde[3]

[1]Universidad de Investigación e Innovación de México (UIIX), Morelos, México https://orcid.org/0000-0001-8662-6415, mcastellanosa@comunidad.uiix.edu.mx
[2]Facultad de Ingenierías, Universidad Tecnológica de Pereira, Pereira, Risaralda, Colombia ORCID: https://orcid.org/0000-0002-9414-6187, lemunozg@utp.edu.co
[3]Universidad Pedagógica y Tecnológica de Colombia (UPTC), Tunja, Boyacá, Colombia PhD en Ciencias Económicas y Empresariales, Universidad de Granada, España Magíster en Administración de Organizaciones, https://orcid.org/0000-0003-3942-8912, efren.baez@uptc.edu.co

Abstract
Artificial intelligence (AI) literacy and AI-related digital competencies are becoming increasingly important in U.S. K–12 education as AI tools expand across instructional, administrative, and professional contexts. In response, teacher professional development (PD) has emerged as a critical area for preparing educators to engage with AI in pedagogically meaningful and ethically responsible ways. This study adopted a hybrid design combining a PRISMA 2020-guided systematic literature review with bibliometric mapping in VOSviewer to examine research on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development published between 2019 and 2026. Records were retrieved from Google Scholar through Publish or Perish.
The search yielded 451 records, of which 73 duplicates were removed in the review workflow. After title and abstract screening of 378 records, 261 were excluded. Full-text assessment of the remaining 117 records resulted in 107 exclusions and a final corpus of 10 included studies. Complementary bibliometric mapping of the filtered topic-relevant dataset revealed a relatively coherent field organized around three interrelated dimensions: AI literacy and teacher-oriented implementation; AI competence and digital competence as the conceptual core; and curriculum design and competency frameworks as the formalization layer. The most visible thematic patterns emphasized classroom practice, teacher preparation, programmatic initiatives, and framework development.
Methodologically, the literature was dominated by exploratory, descriptive, and self-report-based approaches, including surveys, case studies, mixed-methods designs, and validation-oriented studies. A major finding was the absence of strong evidence for objective biometric or physiological assessment. Terms related to eye-tracking, electroencephalography (EEG), heart rate variability, and wearable sensing were not prominent in the bibliometric structure, suggesting that this dimension remains underdeveloped. Overall, the field is expanding but remains conceptually evolving and methodologically constrained, highlighting the need for clearer definitions and more rigorous multidimensional approaches to assessing teacher learning and AI-related competency development.

Keywords: AI literacy; AI-related digital competencies; teacher professional development; U.S. K–12 education; systematic review; bibliometric analysis; biometric assessment

INTRODUCTION

Artificial intelligence (AI) is increasingly reshaping educational practice, policy, and professional learning across the United States. As AI tools become more visible in K–12 settings, teachers are expected not only to use them effectively, but also to understand their pedagogical value, ethical implications, and potential risks. This shift has intensified the need for AI literacy and AI-related digital competencies among educators, particularly in school environments where technology-related decisions directly influence instruction, inclusion, and student opportunity.

AI literacy is generally understood as the set of knowledge, skills, and critical understandings required to interpret, evaluate, and engage with AI systems in informed and responsible ways. In educational settings, this extends beyond basic technical familiarity and includes the ability to recognize how AI tools function, interpret outputs critically, identify limitations and bias, and apply such tools in pedagogically meaningful ways. For teachers, these capabilities intersect with broader forms of digital competence, yet they also require distinct forms of professional judgment related to algorithmic systems, ethical decision-making, and reflective instructional practice in AI-supported environments.

The growing relevance of AI in schools has increased interest in teacher professional development (PD) as a key mechanism for preparing educators to work effectively in AI-rich educational contexts. Professional development in this area may include training on foundational AI concepts, classroom applications, ethical considerations, instructional design, and the responsible use of digital tools. However, the literature on this topic remains fragmented. Some studies explicitly focus on AI literacy, while others use related constructs such as AI competency, digital readiness, teacher preparedness, or technology integration. This conceptual diversity makes it difficult to determine how the field is evolving and how teacher learning in AI-related professional development is being defined, structured, and assessed.

This issue is particularly important in the United States, where K–12 education is shaped by substantial variation across states, districts, and school communities. Differences in infrastructure, staffing, funding, and access to digital resources influence how professional development is designed and implemented. As a result, the development of AI-related teacher competencies cannot be fully understood without considering the complexity of the broader educational landscape in which such learning takes place.

At the same time, questions of assessment remain central to the field. Much of the emerging literature appears to rely on subjective methods such as surveys, interviews, reflection logs, and perception-based instruments. These approaches are useful for capturing teacher attitudes, confidence, and self-reported preparedness, but they may not fully reveal how teachers engage cognitively or behaviorally during professional development. This creates an important methodological concern: if AI-related teacher competencies are becoming increasingly significant, the field requires stronger and more multidimensional ways of evaluating them.

One underexplored possibility is the use of objective biometric and physiological measures. In other areas of the learning sciences and human-computer interaction, tools such as eye-

tracking, electroencephalography (EEG), heart rate variability, and wearable sensors have been used to examine attention, cognitive load, stress, and real-time engagement. However, such approaches appear to be largely absent from the literature on AI-focused teacher professional development. Understanding the extent of this gap is important for the future development of more rigorous and evidence-based approaches to teacher competency assessment.

Given these developments, a study that combines systematic review and bibliometric analysis is timely. A systematic review enables the identification and synthesis of concepts, methods, and findings across the literature, while bibliometric analysis helps map publication trends, thematic clusters, and conceptual relationships within an emerging field. Together, these approaches provide a stronger basis for understanding how research on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development has developed between 2019 and 2026.

Accordingly, this study examines the literature on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development through a systematic review guided by PRISMA 2020 and a bibliometric analysis using VOSviewer. Drawing on records retrieved through Google Scholar and screened through a structured review process, the study investigates publication patterns, conceptual orientations, methodological tendencies, and assessment approaches in the field. Particular attention is given to the extent to which the literature incorporates, or fails to incorporate, objective biometric evaluation in the assessment of teacher learning and competency development

Questions

This study was guided by the following research questions:

RQ1. What are the main publication trends in the literature on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development between 2019 and 2026?

RQ2. How are AI literacy and AI-related digital competencies conceptualized in the reviewed studies?

RQ3. What methodological approaches are most commonly used in studies on AI-focused teacher professional development in the U.S. K–12 context?

RQ4. What themes, keywords, and research clusters emerge from the bibliometric analysis of the selected literature?

RQ5. To what extent do the reviewed studies incorporate objective or biometric evaluation methods to assess teacher learning, engagement, or competency development during professional development?

RQ6. What major gaps and future directions can be identified in the literature, particularly regarding rigorous and multidimensional assessment approaches in AI-related teacher professional development?

METHODOLOGY

This study employed a hybrid review design that combined a systematic literature review with a bibliometric analysis to examine research on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development published between 2019 and 2026. The systematic component was guided by the PRISMA 2020 framework, while the bibliometric

component was used to identify publication patterns, thematic structures, and conceptual relationships within the literature.

The review was designed to address both descriptive and analytical purposes. First, the systematic review synthesized the conceptual, methodological, and empirical characteristics of studies related to AI-focused teacher professional development in the United States. Second, the bibliometric analysis identified patterns in scientific production, including publication growth, source distribution, keyword occurrence, and thematic clustering. Together, these approaches provided both interpretive depth and structural insight into the development of the field.

The primary search source for this study was Google Scholar. This platform was selected because of its broad coverage of scholarly materials across journals, conference proceedings, dissertations, preprints, and other academic outputs, as recognized in prior bibliometric and informetric research (Halevi et al., 2017; Martín-Martín et al., 2018). Because Google Scholar does not provide the same level of indexing control, metadata standardization, and field-specific search functions as curated subscription databases, the review relied on a structured multi-query search process and careful manual screening to ensure relevance and consistency (Halevi et al., 2017; Orduna-Malea et al., 2015).

Search retrieval, record export, and initial citation management were supported through Publish or Perish (Harzing, 2007), which enabled the extraction and organization of Google Scholar results for subsequent screening and bibliometric preparation. This step was especially useful given the broader and less standardized structure of Google Scholar records.

Studies were included if they were published between 2019 and 2026, written in English, focused on U.S. K–12 education, addressed teacher professional development or related forms of in-service teacher learning, and examined AI literacy, artificial intelligence literacy, or AI-related digital competencies in educational contexts. The review considered journal articles, conference proceedings, preprints, and doctoral dissertations when they were directly relevant to the review scope and provided sufficient methodological and conceptual detail for analysis. Studies were excluded if they focused exclusively on higher education, pre-service teacher preparation without a clear in-service K–12 dimension, general digital competence without explicit reference to AI, or non-U.S. contexts.

After search retrieval, records were compiled into a single dataset and screened for duplicates. The remaining studies were reviewed through title and abstract screening, followed by full-text eligibility assessment. Reasons for exclusion at the full-text stage were documented to maintain transparency in the selection process.

A structured data extraction form was used to record bibliographic, conceptual, and methodological information from each included study. Extracted variables included publication year, title, source, educational context, participant group, professional development focus, conceptual orientation, research design, data collection methods, and assessment approaches. Particular attention was given to whether studies used self-report, observational, performance-based, or objective biometric measures to evaluate teacher learning or competency development.

The bibliometric analysis was conducted using VOSviewer. To preserve analytical breadth, the bibliometric mapping was performed on a filtered topic-relevant dataset prepared for visualization, whereas the systematic synthesis focused on the final set of included studies. This distinction allowed the study to map the broader conceptual structure of the field while

still maintaining a clearly defined corpus for systematic review and eligibility-based analysis. Keyword co-occurrence analysis was used to identify recurring concepts and to visualize the relationships among themes such as AI literacy, digital competence, teacher training, ethics, and technology integration.

A central analytical focus of this review was the identification of gaps in objective biometric evaluation. For the purposes of this study, biometric evaluation refers to physiological or sensor-based measures capable of capturing real-time cognitive, affective, or behavioral responses during professional learning processes. These may include eye-tracking, electroencephalography (EEG), heart rate variability, or wearable sensors. The review therefore examined whether such approaches were present, limited, or absent in the selected literature.

The final reporting structure followed PRISMA 2020 and was designed to include a PRISMA flow diagram, summary tables of included studies, and bibliometric visualizations. This methodological approach supported transparency and replicability while also acknowledging the broader and less standardized indexing characteristics of Google Scholar.

PRISMA Search Strategy

The search strategy for this review was developed in accordance with the PRISMA 2020 guidelines to ensure a transparent and replicable identification of relevant studies. The review focused on literature related to AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development published between 2019 and 2026. Records were retrieved through Google Scholar.

Because Google Scholar does not support the same field-based search structure available in subscription databases, the search process was conducted using multiple phrase-based queries rather than a single complex database string. The strategy was organized around four conceptual components: AI literacy and related competencies, teacher professional development, K–12 educational settings, and the United States context.

Searches were limited to the period from 2019 to 2026 using the custom date range feature available in Google Scholar. English-language relevance was determined during the manual screening stage. Records retrieved from all searches were compiled into a single dataset, after which duplicates were identified and removed prior to title and abstract screening. The following search queries were used in Google Scholar:

Search 1: “AI literacy” “teacher professional development” “United States” “K-12”
Search 2: “artificial intelligence literacy” “teacher training” “United States” “K-12”
Search 3: “AI competency” teacher “professional learning” “United States”
Search 4: “AI-related digital competencies” teacher “K-12” USA
Search 5: “artificial intelligence competency” teacher “professional development” “elementary education” “United States”
Search 6: “artificial intelligence competence” teacher “secondary education” USA

To improve search coverage, additional related terms such as educator, school teacher, in-service teacher education, primary education, and secondary education were also considered where necessary. This flexible approach was used to account for terminological variation across the literature.

Although the review title highlights gaps in objective biometric evaluation, biometric terms were not used as mandatory search restrictions. This decision was made to avoid excluding

potentially relevant studies on AI-focused teacher professional development that did not explicitly mention biometric or physiological methods in their titles or abstracts. Instead, the presence or absence of such methods was examined during the full-text review and data extraction stages.

Following PRISMA 2020, the study selection process involved four stages: identification, screening, eligibility, and inclusion. First, all records retrieved through Google Scholar were collected. Second, duplicate records were removed. Third, titles and abstracts were screened for relevance. Finally, the full texts of potentially eligible studies were assessed against the predefined inclusion and exclusion criteria. Studies that did not focus on U.S. K–12 teacher professional development, did not address AI literacy or AI-related digital competencies, or lacked sufficient relevance to the review scope were excluded.

Inclusion and Exclusion Criteria

The eligibility criteria for this review were established in advance to ensure consistency, transparency, and alignment with the study objectives.

Inclusion Criteria

Studies were included if they met all of the following conditions:

1. They were published between 2019 and 2026.
2. They were retrieved through Google Scholar.
3. They were written in English.
4. They focused on U.S. K–12 education.
5. They examined teacher professional development, teacher training, in-service teacher education, or related forms of educator learning.
6. They addressed AI literacy, artificial intelligence literacy, or AI-related digital competencies.
7. They provided sufficient academic and methodological information to support systematic review and bibliometric analysis.
8. They were published as journal articles, conference proceedings, preprints, or doctoral dissertations directly relevant to the review scope.

Exclusion Criteria

Studies were excluded if they met any of the following conditions:

1. They focused exclusively on higher education or university faculty.
2. They addressed pre-service teacher education without a clear in-service K–12 professional development component.
3. They focused only on general digital competence without an explicit AI-related dimension.
4. They were conducted outside the United States or did not clearly identify a U.S. educational context.
5. They were editorials, opinion papers, book reviews, or reports lacking sufficient relevance or methodological detail.
6. Their full text was unavailable for eligibility assessment.
7. They were duplicate records retrieved through multiple Google Scholar searches.

The use of objective biometric evaluation was not a requirement for inclusion. Studies were included even when they relied exclusively on self-report, interview-based, or observational

methods, since one of the aims of the review was to determine the extent to which biometric approaches were present, limited, or absent in the field.

Data Extraction and Coding Process

A structured data extraction and coding process was used to ensure consistency across the review and to support both the systematic and bibliometric components of the study. After the final set of eligible studies was identified, each article was examined in full text and coded using a predefined extraction framework aligned with the research questions and study objectives.

For each included study, bibliographic information was recorded, including author name(s), publication year, title, source, and document type. Contextual information was also extracted, such as the educational setting, teacher population, and the type of professional development addressed. In addition, conceptual information was collected to identify whether the study focused on AI literacy, artificial intelligence literacy, AI competency, or AI-related digital competencies.

Methodological information was extracted from each study to compare research designs and assessment approaches across the literature. This included the research design, sample characteristics, data collection methods, and the type of instruments used to evaluate teacher learning, preparedness, or competency development. Particular attention was given to whether studies used self-report measures, interviews, observations, performance-based tasks, or objective measures.

Because the title of this review highlights gaps in objective biometric evaluation, this dimension received specific attention during coding. Each study was examined to determine whether it included any biometric, physiological, or sensor-based measure, such as eye-tracking, electroencephalography (EEG), heart rate variability, or wearable sensors. Studies were also coded according to whether their assessment approach was primarily subjective, objective, or mixed.

The extracted data were organized into a structured review matrix to facilitate comparison across studies. This matrix supported the identification of recurring concepts, methodological trends, thematic patterns, and evaluation practices. It also provided the basis for the narrative synthesis and the preparation of bibliometric data for analysis in VOSviewer.

To promote consistency, the same coding procedure was applied to all included studies. This stepwise approach helped ensure that each article was reviewed according to the same analytical categories and that the evidence base could be synthesized in a systematic and transparent manner.

Data Analysis

The analysis combined systematic synthesis and bibliometric mapping in order to provide both interpretive depth and structural insight into the literature on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development.

The systematic analysis focused on identifying the main conceptual, contextual, methodological, and evaluative patterns across the included studies. After data extraction and coding were completed, the studies were compared to identify recurring themes, conceptual similarities, methodological differences, and emerging gaps in the literature. Special attention

was given to how studies defined AI literacy and related competencies, how professional development initiatives were described, and how teacher learning outcomes were assessed.

A narrative and thematic approach was used to synthesize the findings of the selected studies. Studies were grouped according to shared conceptual orientations, research designs, and assessment methods. This process made it possible to identify dominant trends in the literature, such as the prevalence of teacher readiness, ethical AI use, and the frequent reliance on self-report and perception-based measures.

The bibliometric analysis was conducted using VOSviewer and focused on a filtered topic-relevant dataset prepared specifically for mapping. This component was used to examine the structural development of the field through the analysis of publication trends, keyword occurrence, and thematic clustering. Keyword co-occurrence analysis was performed to identify the most frequent concepts appearing across the literature and to visualize relationships among terms such as AI literacy, digital competence, teacher training, ethics, professional development, and technology integration.

In addition, the analysis examined publication growth over time, allowing the study to identify whether interest in AI-related teacher professional development increased during the period from 2019 to 2026. The bibliometric component also supported the identification of dominant thematic clusters and the conceptual organization of the field.

A specific analytical layer focused on assessment approaches. Each included study was classified according to whether it used primarily subjective, objective, or mixed methods of evaluation. Subjective approaches included surveys, interviews, reflection logs, and perception-based instruments. Objective approaches included direct performance tasks, observational protocols, analytics, and biometric or physiological tools. Mixed approaches combined two or more of these methods.

Because this review specifically highlights gaps in objective biometric evaluation, the analysis also examined whether the selected studies incorporated physiological or sensor-based methods such as eye-tracking, electroencephalography (EEG), heart rate variability, or wearable sensors. The presence, absence, or limited use of these methods was treated as an important analytical finding rather than as a secondary observation.

The final stage of analysis involved integrating the results of the systematic and bibliometric components. This combined approach made it possible to identify both the substantive development of the literature and its methodological limitations. It also provided the basis for interpreting how the field has evolved and where future research may be needed, particularly in relation to more rigorous and multidimensional approaches to evaluating teacher professional development in AI-related contexts.

Limitations

This study has several limitations. First, the review relied on Google Scholar as the primary search source. Although Google Scholar offers broad scholarly coverage and can retrieve materials that may not always be captured in more selective databases, its indexing logic, database size, and metadata consistency are less transparent and standardized than those of curated subscription-based sources (Halevi et al., 2017; Martín-Martín et al., 2018; Orduna-Malea et al., 2015). As a result, the search process required substantial manual screening and careful duplicate management.

Second, the use of Publish or Perish facilitated retrieval and export from Google Scholar, but the quality of the bibliographic data remained dependent on the structure and completeness of the source records themselves (Harzing, 2007). This means that inconsistencies in titles, author names, source information, or keyword metadata may have affected both the screening process and the bibliometric visualization.

Third, because the field is still emerging and conceptually diverse, relevant studies may use terminology not fully captured by the selected search queries. Variations across terms such as AI literacy, AI competence, digital competence, teacher readiness, and technology integration may have influenced retrieval sensitivity.

Fourth, the study combined two related but not identical analytical corpora: the final included studies used for systematic synthesis and a filtered topic-relevant dataset used for bibliometric mapping. This approach strengthened thematic visualization but should be considered when interpreting the relationship between the systematic review findings and the broader conceptual structure shown in the VOSviewer maps.

These limitations should be considered when interpreting the scope and conclusions of the review.

RESULTS

Study Selection

The study selection process followed the PRISMA 2020 framework. Across six structured Google Scholar searches conducted through Publish or Perish, a total of 451 records were initially identified. For the purposes of the systematic review workflow, duplicate removal at the screening stage resulted in 378 records, indicating the exclusion of 73 duplicate records prior to title and abstract screening. After screening, 261 records were excluded for not meeting the eligibility criteria. The remaining 117 records were assessed at the full-text level, resulting in the exclusion of 107 studies and a final sample of 10 included studies.

Two studies initially placed in the *Maybe* queue were retained after final assessment: MB-89, Understanding teacher perspectives and experiences after deployment of AI literacy curriculum in middle-school classrooms, and MB-94, Investigating changes in teachers' perceptions about artificial intelligence after virtual professional development. Their inclusion strengthened the final review corpus by contributing direct evidence on classroom implementation and in-service teacher learning following professional development interventions.

Overall, the PRISMA process illustrates a substantial narrowing from broad retrieval to focused inclusion. This reduction reflects the breadth and relatively low specificity of Google Scholar retrieval, which yielded many records outside the intended review scope, particularly studies centered on higher education, preservice teacher education, non-U.S. contexts, or general digital competence without an explicit AI-related focus. Consequently, the final sample of 10 studies represents a highly targeted body of evidence aligned with AI literacy, AI-related digital competencies, and U.S. K–12 teacher professional development.

Bibliometric Dataset Preparation

To complement the systematic review, bibliometric mapping was conducted in VOSviewer using a filtered topic-relevant dataset, rather than the final set of 10 included studies. This

distinction is methodologically important, as a corpus of only 10 studies would be insufficient for producing an interpretable co-authorship structure and would substantially limit the value of network-based analysis.

For the bibliometric mapping workflow, the RIS dataset was cleaned separately to prepare the file for VOSviewer. In this preprocessing stage, 71 duplicate records were removed from the exported RIS file, reflecting a more conservative duplicate-detection procedure applied to the bibliometric dataset. Accordingly, the difference between the number of duplicates removed in the PRISMA workflow (73) and in the VOSviewer preprocessing workflow (71) is due to the use of two distinct cleaning procedures for two different analytical purposes. The PRISMA count refers to duplicate removal for study selection, whereas the VOSviewer count refers to conservative deduplication of the RIS file used for bibliometric mapping.

The bibliometric results reported below therefore describe the broader topic-relevant literature captured in the filtered dataset, whereas the systematic review findings refer exclusively to the final corpus of 10 included studies.

Bibliometric Structure of the Field

The VOSviewer co-occurrence map suggests that the field is not fragmented into disconnected lines of inquiry, but instead displays a relatively coherent conceptual organization. The most prominent terms in the network were practice (22 occurrences), classroom (20), initiative (15), program (14), survey (10), theory (10), and ai competency framework (10). Other salient terms included curriculum design (8), literacy education (8), digital competence (6), and digital competency (5).

Taken together, these patterns indicate that the topic-relevant literature develops through interconnections among practical educational implementation, conceptual framing, and curricular formalization, rather than through isolated thematic strands.

Thematic Development

Three major thematic nuclei emerged from the co-occurrence analysis.

The first thematic cluster is associated with AI literacy and teacher-oriented educational implementation. Terms such as *ai literacy education*, *literacy education*, *program*, *classroom*, *context*, and *implication* point to a body of literature concerned with teacher preparation, classroom integration, and the pedagogical consequences of introducing AI into school environments. The visibility of terms such as *case study*, *survey*, and *initiative* suggests that this line of work is predominantly applied, practice-oriented, and exploratory in methodological orientation.

The second thematic cluster is organized around artificial intelligence competence, digital competence, and digital competency. These concepts occupy a structurally intermediate position in the network, indicating that they function as the conceptual core of the field. Their centrality suggests that AI-related teacher capability is frequently interpreted through pre-existing digital competence traditions, rather than through a fully autonomous and stabilized conceptual framework specific to AI.

The third thematic cluster is centered on curriculum design, higher education, theory, practice, course, and ai competency framework. This grouping reflects a line of scholarship oriented toward the formalization of AI-related competencies through curricular design, course development, and framework construction. At the same time, the presence of *higher education*

in the map indicates that the broader mapped dataset extends beyond the final review scope, which helps explain the large number of exclusions during screening.

Bridging Concepts Across the Network

Beyond thematic clustering, the map reveals several connector terms that organize the field relationally. In particular, digital competence and digital competency link classroom-oriented and pedagogical discussions with more formalized, framework-based approaches. Similarly, practice and course connect curricular design to competency modeling, while program, classroom, and context connect AI literacy discourse with educational implementation.

These linkages suggest that the field is developing as a continuum extending from classroom practice and teacher learning to curricular structuring and competency formalization. Rather than constituting separate debates, these strands appear to form an interconnected research agenda.

Methodological Tendencies

The co-occurrence structure also points to recognizable methodological tendencies within the mapped literature. The prominence of terms such as survey, case study, and mixed methods study indicates that the field remains dominated by exploratory, descriptive, and perception-oriented research designs. This pattern suggests a strong dependence on self-report instruments and experience-based accounts, with comparatively limited emphasis on observational, behavioral, or performance-based forms of assessment.

Notably, no biometric or physiological terms emerged as salient elements of the visible network. The absence of terms related to measures such as eye-tracking, electroencephalography, heart rate variability, or wearable sensing reinforces one of the central analytical conclusions of the review: although AI-related teacher competence is increasingly discussed conceptually and pedagogically, the literature remains underdeveloped with respect to objective biometric evaluation.

Author Collaboration Structure

The co-authorship map reveals a limited but discernible collaboration structure within the filtered topic-relevant dataset. Two small collaboration clusters are visible. The first comprises Isaacs, S., Hinostroza, J. E., Senne, F., Orr, D., and Lee, M. The second includes Zhao, X., Chen, X., and Xie, H. Between these two groups, Miao, F. occupies a central bridging position, linking the otherwise separate collaboration clusters.

Overall, the co-authorship structure suggests that the field remains collaborative only at a modest scale. Rather than forming a dense and highly integrated scholarly network, research activity appears to be organized around relatively small teams connected through a limited number of bridging authors. This pattern is consistent with an emerging field in which collaboration is present but not yet consolidated into a broader, mature research community.

Summary of Main Findings

Taken together, the results indicate that the literature on AI literacy and AI-related digital competencies in teacher education is structured around three interrelated dimensions: (1) AI literacy and teacher-oriented implementation, (2) AI competence and digital competence as the conceptual core, and (3) curriculum design and competency frameworks as the

formalization layer. At the same time, the field remains methodologically constrained, with a strong emphasis on exploratory and self-report-based approaches and little visible evidence of objective biometric assessment.

Importantly, these bibliometric patterns derive from the filtered topic-relevant dataset used for VOSviewer mapping, whereas the systematic review itself was based on the final corpus of 10 included studies. The bibliometric analysis therefore provides a broader account of the intellectual and collaborative structure of the field, while the systematic review offers a more selective synthesis of the evidence most directly aligned with the review question.

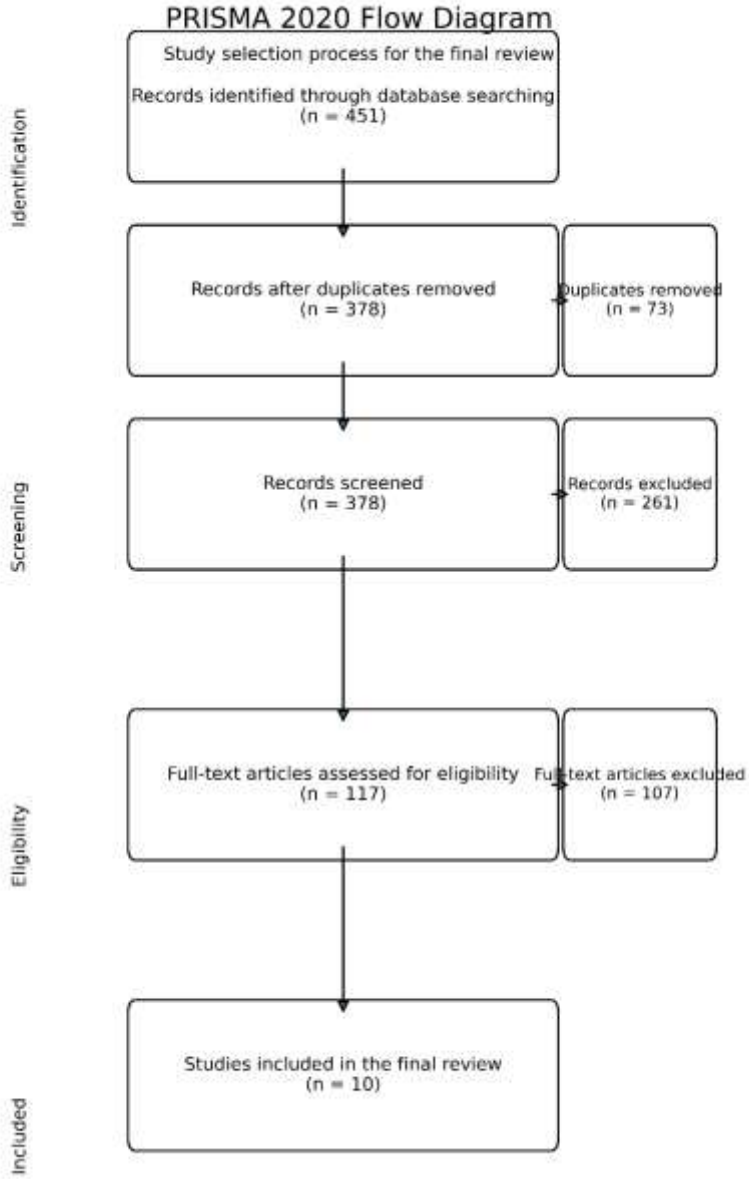


Figure 1. PRISMA 2020 flow diagram of the study selection process.

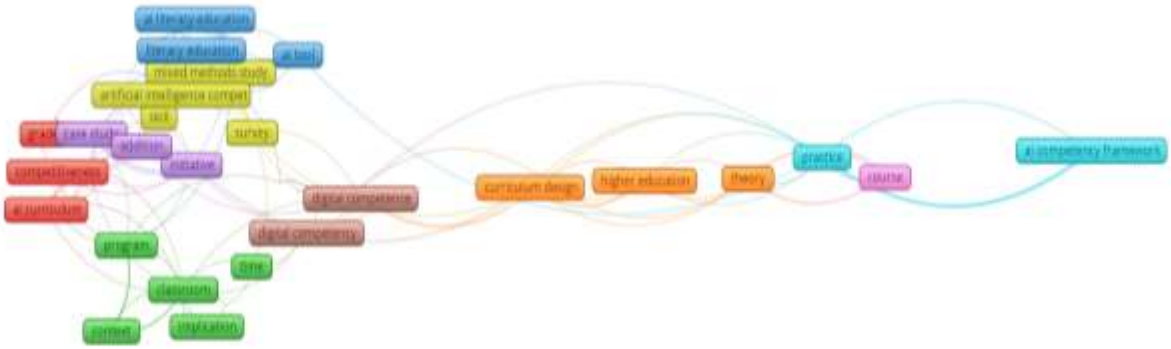


Figure 2. VOSviewer term co-occurrence map of the filtered topic-relevant dataset on AI literacy and AI-related digital competencies in education

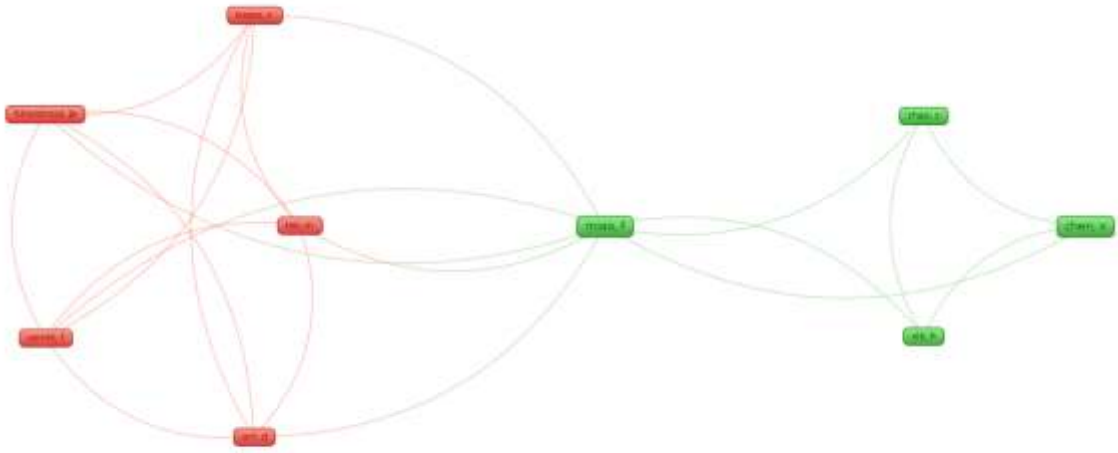


Figure 3 presents the co-authorship network generated in VOSviewer from the filtered topic-relevant dataset. Overall, the network shows

DISCUSSION

The findings of this review indicate that research on AI literacy and AI-related digital competencies in teacher education is gaining visibility, yet remains conceptually uneven and methodologically underdeveloped. Although the initial search retrieved 451 records, only 10 studies met the final eligibility criteria after screening and full-text assessment. This result suggests that, despite the growing presence of AI-related topics in educational research, only

a limited proportion of the available literature addresses the specific intersection of U.S. K–12 in-service teacher professional development and explicit AI-related competency development. Much of the broader literature appears to remain distributed across adjacent domains, including higher education, preservice teacher education, and general digital competence.

The bibliometric analysis helps clarify this broader landscape. The VOSviewer maps, based on the filtered topic-relevant dataset rather than the final review corpus, reveal that the field is already developing a recognizable conceptual structure. Three interrelated thematic dimensions emerged: AI literacy and teacher-oriented implementation, AI competence and digital competence as the conceptual core, and curriculum design and competency frameworks as the formalization layer. These dimensions suggest that the field is not developing through isolated lines of inquiry, but rather through an interconnected agenda linking classroom practice, teacher learning, conceptual framing, and curricular development. At the same time, the coexistence of these strands indicates that the field has not yet stabilized around a unified terminology or a fully consolidated theoretical framework.

One of the most significant findings concerns the central role of digital competence within the thematic network. Its intermediate and connective position suggests that AI-related teacher capability continues to be interpreted largely through pre-existing educational technology frameworks rather than as a distinct professional domain. This continuity is understandable, since digital competence provides an established language for discussing teacher readiness, technology integration, and pedagogical use. However, it may also constrain conceptual development if AI-related competencies continue to be framed merely as an extension of general digital skills. Educational engagement with AI may require more specific constructs related to algorithmic awareness, critical data interpretation, bias recognition, ethical judgment, and pedagogical decision-making in AI-mediated environments. In this sense, the field appears to be in a transitional phase, drawing on prior competence traditions while gradually moving toward a more differentiated conceptual identity.

The results also show that the field remains strongly oriented toward applied and exploratory research. The prominence of terms such as survey, case study, and mixed methods study, together with the patterns observed in the included studies, indicates that much of the literature has focused on teacher perceptions, pilot interventions, and early implementation experiences. This tendency is typical of emerging fields, where exploratory designs play an important role in clarifying concepts, identifying relevant variables, and testing initial training models. Nevertheless, the persistence of these methodological patterns also suggests that the field has not yet progressed substantially toward more robust forms of direct competency assessment or stronger evidence regarding the effectiveness of professional development interventions.

This limitation is particularly important in relation to the focus of the present review on objective biometric evaluation. Across the bibliometric structure, no biometric or physiological terms emerged as prominent elements in the co-occurrence network, and the final review corpus likewise showed little evidence of objective physiological or real-time behavioral assessment. Instead, the literature remains heavily dependent on self-report instruments and perception-based approaches. Although such methods are valuable for capturing beliefs, confidence, and perceived readiness, they provide only a partial account of competence. They do not directly show how teachers process AI-related information under authentic instructional conditions, how they respond to uncertainty or ethical dilemmas, or

how attention, cognitive effort, and engagement fluctuate during professional development activities. As a result, the current evidence base appears stronger in documenting what teachers report than in explaining how AI-related competencies are enacted in practice.

The absence of biometric approaches therefore represents more than a technical omission; it reflects a broader limitation in how teacher professional development outcomes are conceptualized and measured. In related areas of the learning sciences, methods such as eye-tracking, electroencephalography, heart rate variability, and wearable sensing have been used to investigate attention, engagement, cognitive load, and stress regulation during learning processes. Their absence in the present field suggests that research on AI-focused teacher professional development remains at an early stage of methodological maturation. Incorporating such approaches would not require replacing qualitative or survey-based methods, but rather complementing them with more fine-grained and multidimensional evidence about how teachers experience and develop AI-related competencies.

The review also offers insight into the problem of scope and boundary formation in this area of research. The large number of excluded records indicates that broad searches on AI competence and education retrieve a substantial amount of literature that does not correspond directly to the specific population and context examined here. This is itself a relevant finding, as it suggests that the apparent growth of AI-related educational research may overstate the maturity of evidence specifically focused on U.S. K–12 in-service teachers. The visibility of terms such as higher education in the bibliometric map reinforces this interpretation, showing that the broader topic-relevant literature extends well beyond the final review scope. Thus, the field appears considerably broader at the bibliometric level than at the level of directly relevant empirical evidence.

The inclusion of two studies originally located in the final *Maybe* queue further underlines the importance of careful full-text assessment in emerging and terminologically unstable fields. In areas where concepts remain fluid and study titles do not always clearly reflect substantive focus, relevant evidence may easily be overlooked if decisions depend too heavily on title and abstract screening alone. This is especially relevant for AI-related educational research, where terms such as literacy, competence, readiness, digital skills, and professional development are often used inconsistently. The review therefore suggests that future evidence syntheses in this domain may benefit from relatively broad initial screening combined with highly selective final eligibility decisions.

These findings have several implications for future research. First, the field would benefit from greater conceptual precision in distinguishing among AI literacy, AI competence, and general digital competence. Without clearer definitional boundaries, comparisons across studies will remain difficult and framework development will continue to reproduce ambiguity. Second, future studies should move beyond predominantly exploratory and self-report-based designs toward more rigorous and multidimensional approaches, including direct performance assessment, classroom observation, learning analytics, and, where feasible, biometric or physiological measures. Third, more research is needed that specifically addresses the professional learning needs of U.S. K–12 in-service teachers across diverse school settings, where infrastructure, access, institutional support, and policy conditions may vary substantially.

The findings also have practical implications for teacher professional development design. The prominence of classroom- and practice-oriented terms in the bibliometric map is encouraging

because it indicates that the field is not purely theoretical; AI-related professional development is already being discussed in relation to actual educational implementation. At the same time, the stronger visibility of curricular design and framework-oriented work, relative to objective assessment research, suggests a potential imbalance. The field may be advancing more rapidly in defining what teachers should know than in establishing how such learning should be evaluated in valid, reliable, and context-sensitive ways. This gap is important because professional development cannot become genuinely evidence-informed if competency models are elaborated more quickly than the methods needed to assess their impact.

A key limitation of the present study is that the bibliometric analysis was conducted on a filtered topic-relevant dataset broader than the final review corpus. Accordingly, the VOSviewer maps should be interpreted as representing the wider intellectual and collaborative structure of the field rather than only the 10 studies included in the systematic review. This does not weaken the value of the bibliometric findings, but it does require interpretive caution: the network analysis provides a broader contextual view of how the field is organized, whereas the systematic review offers a more selective synthesis of the evidence most directly aligned with the review question.

Overall, the discussion points to a field that is clearly emerging, conceptually active, and increasingly relevant, but still methodologically constrained. If AI literacy is to become a meaningful dimension of teacher professional competence, its development must be assessed not only through self-perceived readiness or abstract knowledge, but also through methods capable of capturing how teachers engage with AI under real pedagogical conditions. Advancing toward more robust, multidimensional, and context-responsive evaluation models will be essential for strengthening both the evidence base and the practical relevance of AI-focused teacher professional development in K–12 education.

Limitations

This review has several limitations that should be acknowledged. First, the final systematic review corpus consisted of only 10 included studies, which limits the breadth of empirical evidence directly available for synthesis. Although this small corpus reflects the narrow and specific scope of the review, it also indicates that the conclusions drawn from the included studies should be interpreted with appropriate caution.

Second, the bibliometric analysis was conducted on a filtered topic-relevant dataset broader than the final review corpus. Accordingly, the VOSviewer maps should be interpreted as representing the wider intellectual and collaborative structure of the field rather than only the 10 studies included in the systematic review. While this broader mapping strategy was methodologically necessary in order to obtain interpretable co-occurrence and co-authorship networks, it introduces an important distinction between the evidence synthesized in the review and the literature represented in the bibliometric analysis.

Third, the search strategy relied on Google Scholar retrieval through Publish or Perish. Although this approach allowed broad coverage, it also introduced challenges related to record heterogeneity, incomplete metadata, duplicate handling, and variation in indexing quality. As reflected in the screening process, a substantial proportion of retrieved studies fell outside the final review scope, particularly those focused on higher education, preservice teacher education, or general digital competence without an explicit AI-related focus.

Fourth, the review was limited to studies aligned with U.S. K–12 in-service teacher professional development, which necessarily excluded relevant work conducted in other national contexts or educational sectors. This narrow scope was appropriate for the research question, but it also means that the findings should not be generalized to teacher education more broadly without caution.

Finally, the review focused specifically on AI literacy, AI-related digital competencies, and biometric or objective assessment approaches. This focus strengthened the analytical specificity of the review, but may also have excluded studies addressing related constructs through different terminology or alternative methodological traditions. Given the terminological instability that still characterizes this field, some relevant evidence may remain difficult to identify through standard search formulations alone.

Implications for Future Research

The findings point to several priorities for future research. First, greater conceptual precision is needed in distinguishing among AI literacy, AI competence, and general digital competence. At present, these concepts often overlap in ways that make cross-study comparison difficult and hinder the development of cumulative knowledge. Future research should aim to clarify the boundaries, dimensions, and practical implications of these constructs, particularly in relation to teacher professional learning.

Second, the field would benefit from a stronger move beyond predominantly exploratory and self-report-based designs. Although surveys, case studies, and mixed-methods approaches have played an important role in the early development of the field, future studies should incorporate more direct and multidimensional forms of assessment. These may include performance-based tasks, classroom observation, scenario-based evaluation, learning analytics, and structured measures of pedagogical decision-making in AI-rich contexts.

Third, the absence of biometric and physiological approaches suggests an important methodological frontier. Future research should examine whether tools such as eye-tracking, electroencephalography, heart rate variability, or wearable sensing can contribute to a more nuanced understanding of teacher learning, attention, engagement, and cognitive effort during AI-related professional development. Such methods should not replace qualitative and survey-based approaches, but rather complement them in order to generate richer and more process-sensitive evidence.

Fourth, more research is needed on U.S. K–12 in-service teachers specifically. Much of the broader literature identified during screening extended into higher education, preservice teacher preparation, or non-U.S. settings. Future studies should therefore focus more directly on the conditions, constraints, and professional learning needs of practicing K–12 teachers, including differences across grade levels, school contexts, subject areas, and resource environments.

Finally, future research should place greater emphasis on implementation and impact. Beyond describing teacher perceptions or validating conceptual frameworks, the field now needs stronger evidence regarding whether AI-related professional development actually changes instructional practice, supports pedagogical judgment, improves critical AI understanding, or enhances student-facing educational outcomes. Without such evidence, the field risks advancing conceptually faster than it advances empirically.

Practical Implications

The findings of this review also carry several practical implications for the design of teacher professional development. First, professional development initiatives should move beyond broad awareness-raising and provide teachers with opportunities to develop competencies that are specifically relevant to AI-rich educational environments. These include not only technical familiarity, but also ethical reasoning, critical interpretation of AI-generated outputs, bias awareness, and pedagogical decision-making.

Second, the strong visibility of classroom- and practice-oriented terms in the bibliometric map suggests that AI-related professional development should remain closely connected to real instructional contexts. Programs are likely to be more meaningful when they are embedded in authentic teaching scenarios, classroom problem-solving, and subject-specific implementation challenges, rather than delivered only as abstract conceptual training.

Third, the findings indicate a need for greater alignment between competency frameworks and assessment strategies. At present, the literature appears to be developing frameworks and curricular models more rapidly than it is developing robust methods to evaluate whether those models actually improve teacher learning or classroom practice. Professional development design should therefore incorporate valid and context-sensitive assessment mechanisms from the outset, rather than treating evaluation as a secondary concern.

Fourth, school leaders, policymakers, and teacher educators should be cautious about assuming that teacher confidence with AI necessarily reflects teacher competence. Because much of the current literature relies on self-report measures, confidence and readiness may be overinterpreted in the absence of more direct evidence of performance. Professional development systems should therefore aim to assess not only what teachers say they know, but also how they apply that knowledge in educational decision-making.

Finally, the broader educational relevance of AI-related professional development depends on its responsiveness to context. Differences in school infrastructure, institutional support, policy guidance, and teacher readiness are likely to shape how AI is understood and implemented in practice. Effective professional development should therefore be adaptable, context-sensitive, and attentive to issues of access, equity, and pedagogical relevance.

CONCLUSION

The present review indicates that research on AI literacy and AI-related digital competencies in U.S. K–12 teacher professional development is emerging as a recognizable but still insufficiently consolidated field. Although the broader literature on AI and education is expanding rapidly, only a limited number of studies aligned closely with the specific review focus, highlighting the scarcity of directly relevant evidence for in-service teachers in the U.S. K–12 context.

The findings show that the field is structured around three interrelated dimensions: AI literacy and teacher-oriented implementation, AI competence and digital competence as the conceptual core, and curriculum design and competency frameworks as the formalization layer. Together, these dimensions suggest that the literature is moving toward greater coherence, but has not yet reached full conceptual stabilization. In particular, the continued centrality of digital competence indicates that AI-related teacher capability is still frequently

framed through pre-existing technology competence models rather than through a distinct and fully differentiated conceptual framework.

At the methodological level, the available evidence remains dominated by exploratory, descriptive, and self-report-based approaches. Surveys, case studies, and mixed-methods designs continue to shape much of the field, while more direct and multidimensional forms of competency assessment remain limited. Most notably, the review found little evidence of objective biometric or physiological evaluation in AI-focused teacher professional development. This absence points to a significant methodological gap, since self-reported perceptions alone cannot fully capture how teachers engage with AI-related content, make pedagogical judgments, or experience the cognitive and affective demands of professional learning.

Taken together, these findings suggest that the field is advancing conceptually faster than it is advancing methodologically. Future research should therefore prioritize clearer distinctions among AI literacy, AI competence, and digital competence, while also adopting stronger evaluation designs that combine self-report data with direct performance measures, classroom-based evidence, learning analytics, and, where feasible, biometric or physiological approaches. Such progress will be essential for building a more robust evidence base for AI-related teacher professional development.

In sum, this review highlights both the promise and the current limitations of this emerging area of inquiry. If AI literacy is to become a meaningful and evidence-based dimension of teacher professional competence, research must move beyond broad conceptual discussion and toward more rigorous, context-sensitive, and multidimensional approaches to assessment and implementation.

REFERENCES

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

2. 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., Hróbjartsson, 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

3. Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu.* Publications Office of the European Union. https://doi.org/10.2760/159770

4. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

5. Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation: Review of the literature. *Journal of Informetrics, 11*(3), 823–834. https://doi.org/10.1016/j.joi.2017.06.005

6. Harzing, A.-W. (2007). *Publish or Perish* [Computer software]. Harzing.com. https://harzing.com/resources/publish-or-perish

7. Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160–1177.
8. https://doi.org/10.1016/j.joi.2018.09.002
9. Orduna-Malea, E., Ayllón, J. M., Martín-Martín, A., & Delgado López-Cózar, E. (2015). Methods for estimating the size of Google Scholar. *Scientometrics, 104*(3), 931–949. https://doi.org/10.1007/s11192-015-1614-6
10. Chen, X. (2024). *A phenomenological study of teachers' perception of professional development sessions focused on artificial intelligence: Utilizing the TPACK framework to address the professional development needs of teachers* (Publication No. 31638797) [Doctoral dissertation, Pepperdine University]. *ProQuest Dissertations & Theses Global*.
11. https://search.proquest.com/openview/7edac2f62b3307ae0f453997a37e28d2/1?pq-origsite=gscholar&cbl=18750&diss=y
12. Kim, J. (2025). Perceptions and preparedness of K–12 educators in adopting generative AI. *Research in Learning Technology*. https://journal.alt.ac.uk/index.php/rlt/article/view/3448
13. Kurz, T., Jayasuriya, S., Swisher, K., Mativo, J., Pidaparti, R. & Robinson, D.T. (2022). Investigating Changes in Teachers' Perceptions about Artificial Intelligence after Virtual Professional Development. *Journal of Interactive Learning Research, 33*(4), 225-241. Waynesville, NC: Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/221764/.
14. Li, Z., Wang, C., & Bonk, C. J. (2025). Generative AI for teachers' self-directed professional development: A mixed-methods study. *TechTrends*.
15. https://link.springer.com/article/10.1007/s11528-025-01123-8
16. Mayrell, D. (2025). *Enhancing educator AI literacy for effective use of large language models: Exploring teacher usage and professional development* [Doctoral dissertation, ProQuest Dissertations & Theses Global].
https://search.proquest.com/openview/b3e4b5c04ed28973f5fc8f1d603f826d/1?pq-origsite=gscholar&cbl=18750&diss=y
17. Nelson, K. (2025). *K–12 educators' comfortability and perception of artificial intelligence* [Doctoral dissertation, Auburn University]. https://etd.auburn.edu/handle/10415/9833
18. Ospina Tabares, S., Mathayas, N., Mead, H., Mouza, C., & Pollock, L. (2025). Exploring elementary educators' perceptions of AI in the context of professional development. In R. Jake Cohen (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1924–1933). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/225751/
19. Ravi, P., Broski, A., Stump, G., Abelson, H., Klopfer, E., et al. (2023). Understanding teacher perspectives and experiences after deployment of AI literacy curriculum in middle-school classrooms. https://arxiv.org/abs/2312.04839
20. Unal, A., & Unal, Z. (2024). Evaluating the integration of artificial intelligence (AI) in K–12 education. *Journal of Interactive Learning Research*.
https://www.learntechlib.org/p/224682/