Volume 7, Number 11, November 2026. pp. 4166-4183
E-ISSN:2723 – 6692
P-ISSN:2723– 6595
https://doi.org/10.59141/jiss.v7i11.2609

From AI Literacy to Employee Performance: A Systematic Review of Learning and Training Transfer

Nuju Widodo* , Tasrim
Universitas PGRI Yogyakarta, Indonesia
Email : nuju.widodo@gmail.com, asrilpalopo@gmail.com

Keywords:
employee AI literacy; training transfer; performance; human resource development; systematic literature review.

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Publikasi Indonesia

Abstract

Effective workplace use of artificial intelligence (AI) requires employees to identify appropriate tasks, critically evaluate AI-generated outputs, and exercise professional judgment. However, the extent to which AI-related competencies acquired through training are transferred to everyday workplace practices and contribute to improved employee performance remains unclear. This systematic review examined how employee AI literacy was measured, developed through professional training, and associated with training transfer and job performance. Peer-reviewed empirical studies published between 2019 and October 2026 were identified through searches conducted using two web search engines and supplementary reference screening. Methodological quality was assessed using the 2018 Mixed Methods Appraisal Tool (MMAT). Given the heterogeneity of measurement approaches and research designs, the findings were synthesized thematically. The review included 43 empirical reports, of which 10 evaluated professional training programs. AI literacy was assessed using self-reported measures, objective knowledge tests, and proxy indicators derived from professional experience or employee profiles. Training programs commonly incorporated practical exercises, project-based learning, and reflective activities to develop AI-related competencies. Although several studies reported the application of acquired skills in workplace settings, evidence of sustained training transfer and long-term competency utilization remained limited. Variations in outcome measures and the predominance of nonrandomized research designs constrained the ability to establish whether AI literacy directly improved employee performance. Based on these findings, the review proposed a conceptual framework linking professional training, AI literacy development, training transfer, and job performance. Future research should empirically examine these relationships through longitudinal workplace assessments and independent, objective performance indicators, including task quality, task completion time, and error rates.

INTRODUCTION

The effective use of artificial intelligence (AI) in workplace settings requires employees to identify appropriate tasks, provide clear instructions to AI tools, critically evaluate generated outputs, and uphold professional responsibilities . Interviews with knowledge workers have highlighted the importance of AI literacy in facilitating responsible human–AI interaction and

context-appropriate technology use, while needs assessments conducted in small and medium-sized enterprises (SMEs) have emphasized the relationship between employee competencies and organizational readiness. Accordingly, access to AI tools should be considered alongside employees' technological competencies, professional responsibilities, and organizational working conditions (Kosonen & Saltiola, 2026; Passlack et al., 2026).

In this study, employee AI literacy refers to the competencies required to understand, use, critically evaluate, and responsibly apply AI technologies in workplace tasks. This conceptualization encompasses both conventional and generative AI applications and extends beyond programming or technical expertise (Cetindamar et al., 2024). Existing instruments for assessing employee AI literacy employ various approaches, including self-assessment scales, objective knowledge tests, and proxy indicators of competency. For example, recently developed AI literacy scales for employees distinguish among operational proficiency, prompt formulation, output quality evaluation, innovative applications, and ethical awareness. Variations in measurement instruments, competency dimensions, and operational definitions may influence how the relationships between AI literacy and workplace outcomes are interpreted and compared across studies (Almatrafi et al., 2024; Dolezel et al., 2025; Hutabarat et al., 2025; Liu et al., 2025; Long & Magerko, 2020).

The development of AI-related competencies also raises important questions regarding the transfer of training outcomes to workplace practices. Baldwin and Ford conceptualized training transfer as the generalization of acquired knowledge and skills to the work environment and the maintenance of these competencies over time. Although improvements in assessment scores immediately following training may indicate successful learning, demonstrating the application of acquired competencies across different workplace tasks and their sustained use requires additional evidence. The professional AI literacy training programs examined in previous studies incorporated practical exercises, project-based activities, and reflective learning strategies. However, these programs differed in implementation scope, training intensity, and follow-up duration, limiting the comparability of their outcomes and the assessment of sustained training transfer (Baldwin & Ford, 1988; Kong et al., 2024; van Kooten et al., 2024; Wiggins et al., 2020).

Employee performance must also be conceptualized and measured at an appropriate level of analysis. AI adoption readiness, self-efficacy, employee well-being, work agility, and organizational implementation capabilities represent relevant outcomes of AI integration; however, they are conceptually distinct from direct measures of individual task quality, productivity, and job performance. Experimental research involving consultants has demonstrated that the performance benefits associated with AI access depend on task characteristics and the suitability of AI applications. Nevertheless, existing findings have not conclusively established an integrated causal pathway linking professional AI training, improvements in AI literacy, the transfer of acquired competencies to workplace practices, and subsequent changes in employee performance. This limitation highlights the need for research designs that distinguish learning outcomes from behavioral transfer and objectively measured

job performance (Chandra, 2026; Çoban & Beydağ, 2026; Dell'Acqua et al., 2026; W. Liu et al., 2026; Pinski et al., 2024).

METHOD

Eligibility Criteria

The inclusion criteria encompassed empirical studies involving employees, managers, self-employed individuals, and professionals in workplace settings that measured or explored artificial intelligence (AI) competencies or explicitly evaluated AI-related skills development. Eligible studies were full-text, peer-reviewed journal articles or conference papers published in English or Indonesian between January 2019 and October 12, 2026. Studies involving teachers and healthcare professionals were included when they examined work-related responsibilities or professional competency development. Studies involving students without a professional or workplace context, opinion articles, review papers, conceptual publications lacking empirical data, and simulation-only studies were excluded from the core corpus. Methodological quality was assessed but was not used as an automatic exclusion criterion (Table 1).

Table 1. Operational inclusion and exclusion criteria

Dimensions	Inclusions	Exclusions or restrictions
Population	Workers, managers, self-employed, professionals; in-service teachers and health workers	Students without a work context; mixed samples without separable worker outcomes
Focus	Measuring or exploring AI competencies, or evaluating AI skill development	Common use of AI without operationalization of competencies or skills programs
Types of evidence	Actual empirical data; qualitative, quantitative, or mixed	Simulation data, opinions, reviews as primary studies, concepts without empirical data
Publications	A complete peer-reviewed journal or conference paper	Promotional pages, article sets, or versions that aren't yet eligible
Period	First publication in 2019 until October 12, 2026	Outside the time window
Language	English or Indonesian	Other languages
Data adequacy	Adequate methods and results for extraction	Reports with unconfirmed data are withheld from synthesis

Search Sources and Strategies

The primary literature search was conducted on October 12, 2026, using two web search engines, followed by screening publisher websites, institutional repositories, and conference proceedings, as well as selective reference searching. The search strategy comprised four thematic clusters: AI literacy, training and learning transfer, employee performance, and underlying mechanisms and workplace conditions, with 28, 28, 24, and 24 queries, respectively. An additional 37 queries were conducted to verify citations and publication versions. Search terms combined “AI literacy,” “artificial intelligence literacy,” “AI competence,” and “generative AI skills” with terms related to employees, workers, workplaces, professionals, training, learning transfer, performance, and organizational conditions. The Boolean search blocks served as adaptable search templates rather than standardized queries applied identically

across all sources. Furthermore, the number of web search results was not treated as equivalent to the number of unique studies identified.

A supplementary search update was conducted on October 13, 2026, using two additional queries. Of the 26 records screened, one was identified as a duplicate based on its digital object identifier (DOI). Among the remaining 25 records, 22 were excluded during preliminary screening, and three underwent full-text eligibility assessment. One additional report met the inclusion criteria, while two were excluded. This supplementary search was not intended to constitute a comprehensive citation index search. Table 2 summarizes the search procedures and documents 143 queries, comprising 141 initial discovery and verification queries and two supplementary update queries. These figures represented the number of search queries conducted rather than the number of unique studies identified.

Table 2. Summary of searches performed

Cluster	Query	Dates and sources	Example query
Literacy	28	October 12, 2026; web	"AI literacy" employees workplace empirical study
Training and transfer	28	October 12, 2026; web	"AI literacy" "training transfer" journal
Performance	24	October 12, 2026; web	"AI literacy" "job performance" employees empirical
Mechanism and conditions	24	October 12, 2026; web	"AI literacy" "self-efficacy" employees mediation
Citations and versions	37	October 12, 2026; web, publisher, repository	DOI Search
Limited updates	2	October 13, 2026; web	"generative AI literacy" "training transfer" employees

Selection, Appraisal, Extraction, and Synthesis

The initial register contains 149 curated candidates; the raw total of search engine results is not available. Deduplication checks for DOIs, URLs, and titles, and conference, journal, or repository versions are linked to a single report when the publication identity is the same. Evidence family IDs are used conservatively to mark potential sample relationships; the number of families does not ensure the entire independent study. Two administrative staff course reports from Kong and colleagues (2024, 2025) were retained in a single family and the sample sizes were not summed up. Three reports for which empirical data adequacy has not been confirmed remain visible in the selection flow but do not supply synthesis results. The use of AI aids in decision making and source checking; two independent human assessors, human adjudication, and agreement statistics have not been conducted, so this is recognized as a limitation.

The appraisal used MMAT 2018 according to the design category, with the responses of each Y/N/U item accompanied by the reason and location of the source. Unclear responses were not converted to no, and the total score or composite quality rating was not calculated (Hong et al., 2018). Extraction separated context, sample units, literacy measurements, programs, design and timing, outcomes, implementations, mechanisms, and limitations; tool development, validation, wave, and subset samples were not summed without basis, and objective ability measures were separated from perceptions. Coefficients, confidence intervals, standard errors, and zero and negative results were not extracted evenly across the corpus.

The synthesis organizes results deductively into competencies, learning, application, and outputs, and then compares similarities and differences into six overlapping themes. The heterogeneity of constructs, designs, and statistics means that meta-analyses are not performed; no significance-based voting or formal certainty assessments are used. The sustainability of conclusions is compared across the corpus and three subsets (without critical reports, without proceedings, and neither). This comparison assesses the sustainability of thematic conclusions, not the stability of the magnitude of effects or evidence that the results are bias-free.

RESULTS AND DISCUSSIONS

Selection Flow and Corpus Characteristics

Of the initial registers of 149 candidates and 26 update entries, one duplicate was removed so that 174 records were screened. The initial exclusions totaled 51, leaving 123 reports to search. A total of 64 reports were not obtained in adequate text form and 59 were assessed for feasibility. Thirteen reports were issued after assessment (non-compliant population 4, AI competency or program not compliant 4, conceptual work or review 3, actual non-empirical data 1, publication type not suitable 1), resulting in 46 reports selected in 45 conservative families. Three reports were held on data adequacy checks, resulting in 43 reports in 42 active families for synthesis (Figure 1). This flow follows the number of registers, not the search engine raw total.

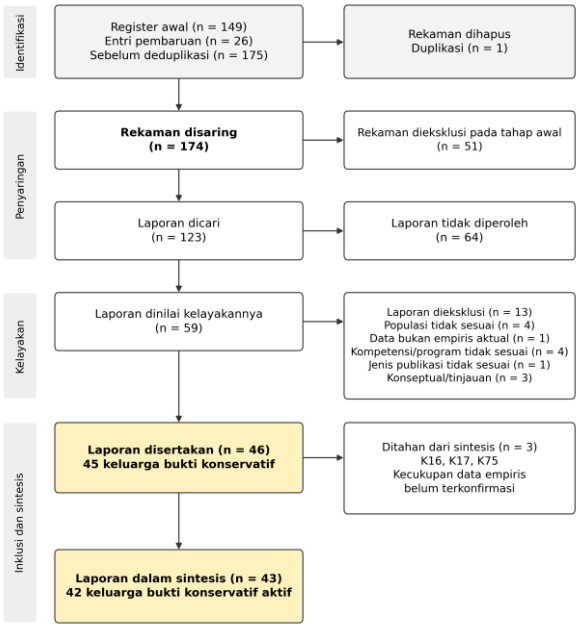


Figure 1. Workflow chart of the selection of employee AI literacy reports (adapted from PRISMA 2020; Page et al., 2021)

A total of 36 reports were journal articles and seven were proceedings. The corpus contains one report in 2020, one in 2023, five in 2024, eleven in 2025, and 25 in 2026 (partial period); this composition describes the selected corpus, not the growth rate of the entire literature. China was the context of 12 reports (27.9%), Indonesia and the United States had four, four cross-regional reports, and four did not yet have a confirmed sample country. The context included corporate workers, administrative staff, teachers, health workers, librarians, leaders, and consultants. Non-random quantitative predominated (23 reports, 53.5%), so the

readings were directed at association and boundary change; two randomized experiments mainly manipulated access or interface, rather than literacy changes due to training (Table 3). The number of participants was not combined because the units of analysis were different.

Table 3. Report design used in synthesis (N = 43)

MMAT Categories	Reports	Implications of reading
Qualitative	5	Explain experiences and needs; without the magnitude of the effect
Randomized experiments	2	Access or interface treatment needs to be distinguished from literacy change
Non-random quantitative	23	Associations, program changes, or path models are not causal automatic
Descriptive	5	Mapping sample competency, readiness, or experience
Mixed method	8	Qualitative, quantitative, and integration components are assessed separately

Definition and Measurement of AI Literacy (RQ1)

Evidence of measurement showed at least three forms of operationalization: self-assessment of ability, concept tests or quizzes, and profile or experience proxies. Liu and colleagues developed generative scales in a sample of 278 exploration, 306 validation, and model testing on 344 public sector workers; all three samples had different functions. The adaptation of SAIL4ALL in retail employees and the professional assessment of health information broadens the context of the measurement, but does not prove the equivalence of the score with other scales or tests (Dolezel et al., 2025; Hutabarat et al., 2025; Liu et al., 2025). Sector comparisons by Lastovska and colleagues and librarian mapping are based on self-assessment, whereas in top management literacy proxies need to be differentiated from actual ability testing. Belief in being able or feeling familiar is not enough to state that one is capable of detecting erroneous outputs, so RQ1 conclusions are constructed heterogeneity and the need to distinguish demonstrated abilities from perceptions (Lastovska et al., 2026; Lo, 2024; Pinski et al., 2024).

Table 4. Comparison of literacy measurement forms and related outcomes

Shape	Sample report	Meaning and limits
Self-assessment scale	K01; T04; K14; K48	Perception of competence or experience; not ensuring the performance of objective tasks
Concept quiz	K39; T26	Assess specific knowledge; Equivalence across instruments has not been tested
Confidence or familiarity	K45; K52; UP15-01	Learning-related perceptual results; not an automatic ability test
Projects and tasks	K51; K46	Work on actual activities during the program; the source of the assessor and the follow-up time need to be differentiated
Executive profile proxies	K79	Indicators of leadership or organizational level; not literacy tests for each worker

Shape	Sample report	Meaning and limits
Use or intention of adoption	K24; K69	Behavior or intention to use; does not measure the entire domain of competence

Source: report extraction. Report code refers to Appendix A. This comparison does not state that one form of measurement is always more valid.

Professional Learning and Programme Outcomes (RQ2)

Professional programs use tool practice, basic concepts, ethics, projects, reflection, and mentoring. Administrative staff courses, teacher development, leader learning, radiographer modules, and data science pathways for residents demonstrate a contextual approach to professional tasks. The evaluation of the radiographer module involved seven participants and six faculty and provided a learning experience, not an estimate of the effect on professional productivity (Alghowinem et al., 2024; Kong et al., 2025; Looi et al., 2026; van de Venter et al., 2023; Wiggins et al., 2020). Table 5 shows examples of results that can be located at the source, not a comparison of effectiveness between packages.

Table 5. Examples of learning outcomes with interpretation limits

Source	Size and yield	Limits
Kong et al., 2024	Concept quiz 4.26 (SD 1.82) to 5.08 (SD 1.62); $p = 0.021$	Courses without random comparators; statistical labels of problematic tables
van Kooten et al., 2024	Retrospective knowledge confidence of 3.25 (SD 1.48) to 6.50 (SD 0.90); $p = 0.002$; 12 pilot participants	Self-assessment; 17 need respondents are different groups
Zhang et al., 2025	Effectiveness of learning perception to skill beliefs: $\beta = 0.084$; $p = 0.122$; $n = 437$	Insignificant; the survey cuts the latitude of the program participants, not the changes resulting from the program

In the co-skilling report, the learning path to skill confidence was not significant, so learning perception was not always followed by application belief; the source claim that all HTMTs below 0.85 also did not match some of the table values, and the report was still included with appraisal restrictions (Zhang et al., 2025). Sebok and colleagues' report used 239 initial responses and 119 final responses with unconfirmed respondent pairs, and the difference in frequency of use written was inconsistent with the two averages shown, so the figure was not used as a verified change. The competency study of psychiatric nurses showed other professional education evaluations, but the nonrandom design and perception measures limited causal attribution. Overall, RQ2 has not yielded conclusions about the best format, duration, or dose (Khalil et al., 2025; Sebok et al., 2025).

Table 6 summarizes ten professional program evaluations. The UP15-01 report is placed on an association orientation because it comes from a survey of co-skilling participants, rather than a standardized pre- and post-program evaluation. Differences in duration, profession, instrument, and comparator prevent the use of tables as a ranking of package effectiveness.

Table 6. Summary of professional program evaluations

Reports	Program and design	Transfer yield and limits
K38 Kong et al., 2025	30 hours five days; professional AI literacy; one group	Acceptance motivation and reflection on implementation; maintenance has not been confirmed; potential overlap with K39
K39 Kong et al., 2024	Three months, 10 sessions × 3 hours; work needs project; before-after without control	Concept test 4.26 to 5.08; p = 0.021; chatbot implementation is self-reported; the table t-value label contains the value of p
K41 Riggs, 2025	AI Summer Institute based on Five Big Ideas in AI; open-ended survey	Classroom perception and practice reported by teachers; N follow-up and duration need clarification
K43 Looi et al., 2026	Six sessions × 3 hours over six weeks; before-after	i-TPACK perception; training plans and tasks; test pairs need clarification
K44 Khalil et al., 2025	AI competency program; quasi-experiment of one group	Knowledge and attitude; clinical performance and maintenance have not been confirmed
K45 Sebok et al., 2025	Adopt four weeks, about 30 minutes per week; early-end surveys	Confidence 3.59 to 3.93; familiarity 2.99 to 3.39; about 45% complete the program; unclear partner
K46 Alghowinem et al., 2024	Four-day experiential workshop; one group	Knowledge and learning experience; ongoing transfer not yet confirmed
K50 van de Venter et al., 2023	Postgraduate AI module; participatory qualitative action research	Learning experience and design; without quantitative effect
K51 Wiggins et al., 2020	Fourth-year Data Science Pathway; immersion and mentoring; descriptive pilot (N = 3)	Actual project involvement during the program; cannot be equated with clinical productivity
K52 van Kooten et al., 2024	Three days × 8 hours; 24-hour curriculum; final survey without control	Confidence 3.25 to 6.5; p = 0.002; groups 17 and 12 were different, not cohort attrition

Source: report extraction. All reports do not have confirmed evidence of maintenance.

Deployment, Maintenance, and Relationship to Performance (RQ3 and RQ4)

Implementation of work appears in participant reports of administrative staff courses, changes in teaching practice, and experiences of participating in generative AI learning. Actual projects on the radiology residency pathway provide evidence of assignment completion during the program, but the source of the assessment, timing, and its relationship to change in abilities differ; project work within the program does not automatically become a postprogram generalization (Kong et al., 2024; Riggs, 2025; Sebok et al., 2025; Wiggins et al., 2020). In the available extracts, the maintenance of competencies and proper use after the program has not been adequately confirmed. This statement does not mean that the article does not have a follow-up, but rather that some details cannot be ascertained, so the RQ3 conclusions are limited to indications of application and the need for repeated observations. Surveys of needs,

frequency of use, or adoption intent are treated as related information, not a substitute for transfer size (Lo, 2024; van Kooten et al., 2024).

Corporate and academic studies examined the relationship between literacy and innovative performance or work behaviors through surveys and pathway models, with task-technology suitability, work engagement, creative efficacy, and pedagogical competence as distinct explanations. The models showed relationships on the analyzed data, but did not estimate the effects of the same training package, and the positive relationships were not combined as universal quantities (Lee & Noh, 2026; Naveed et al., 2026; Tan et al., 2026; Xu et al., 2026). Nepalese industry studies distinguish the use of AI from literacy and report a direct relationship of use with insignificant performance; in Indonesia, digital literacy and competence are associated with hotel worker readiness, while digital leadership and task suitability are examined in the creative industry. Different constructs and designs make readiness and performance not interchangeable (Karn et al., 2026; Novitasari et al., 2025; Setiana et al., 2026; Yang & Lan, 2026). Other related outcomes include nurse thriving, medical AI readiness, academic management innovation, health workforce efficacy, lecturer well-being, employee agility, and organizational AI implementation capabilities; uniform extraction is not yet available to infer the direction and magnitude of the overall relationship between literacy and individual performance (Chandra, 2026; Che, 2026; Çoban & Beydağ, 2026; W. Liu et al., 2026; Y. Liu et al., 2026; Pinski et al., 2024; Xie et al., 2026).

Support, culture, accountability, social norms, and practice opportunities help explain the context of competency use. In interviews, support or norms are the theme of experience, whereas in surveys they can be model variables, so they are not stated as evidence of the same moderation (Fu et al., 2026; Nguyen et al., 2026; Steinhauser & Heid, 2026; Xia et al., 2026; Zhang & Lee, 2025). Autonomy, cognitive dependence, role ambiguity, and identity threat are also examined, but the coefficient of moderation or mediation does not guarantee temporal mechanisms (Li & Mao, 2026; Rabbani et al., 2026; Zhang et al., 2026). Two experiments demonstrated the need to read treatment appropriately: Kalff and Simbeck randomized the explanatory features of the interface for HR managers, while Dell'Acqua and colleagues compared without AI, AI access, and access with introduction to use in 758 consultants, with task-dependent benefits and disadvantages. Differences in access or interfaces are not entirely an effect of literacy improvement and do not prove competency maintenance (Dell'Acqua et al., 2026; Kalff & Simbeck, 2025).

Table 7. Synthesis of themes and answers to research questions

Themes and RQ	Defensible conclusions	Examples of evidence and limits
Measurement (RQ1)	Self-scale, quizzes, tasks, and proxies operationalize different aspects	K01; T04; T26; K79. Size equivalence has not been tested
Learning (RQ2)	Practices, projects, reflections, and mentoring appear in the program	K38; K39; K43; K50; K51; K52. The best format has not yet been determined
Implementation (RQ3)	There is an indication of the use of work and projects during the program	K39; K41; K45; K51. Post-program maintenance has not been adequately confirmed
Work (RQ4)	Performance, readiness, efficacy, well-being, and organizational level are differentiated	K20; K22; K27; K62; K64; K66; K79. No amount of combined effect

Themes and RQ	Defensible conclusions	Examples of evidence and limits
Organizational conditions (RQ4 and RQ5)	Task suitability, support, norms, and autonomy help clarify the context	K03; K04; K10; T01; K63; UP15-01. Experience themes and moderation variables are differentiated
Access treatment (RQ2 and RQ4)	Tool access and meaningful interface changes are different from training	T02; D21. The effect of access is not entirely the effect of literacy change

Source: thematic synthesis. Membership of the theme is not exclusive; frequency indicates the location of the evidence, not the strength of the relationship.

Appraisal and Subset Comparison

The assessment resulted in 217 Y, 148 U, and 22 N responses on 387 items, including screening items. This frequency indicates the completeness of the appraisal responses, not the corpus quality score; the number of unclear answers indicates the need for information verification without proving that the procedure was not performed. Number inconsistencies were noted, and the affected numbers were not used for the combined estimate.

Table 8. Substantive comparison of conclusions on a subset

Subset	Reports	Interpretation changes
The entire corpus	43	Six themes; sample implementation and interface experiments are available
Without six critical reports	37	Size mapping and survival results; conditions of the issued reports are not generalized
Without seven proceedings	36	Examples of quizzes or staff programs and HR interfaces are reduced or missing; implementation is still supported by radiology faculty and projects
Without both	31	Differences in construct and contrast of access in the persistent consultant experiment; the continuous transfer pathway remains uncertain

When the proceedings are issued, the conclusions about the HR interface experiment no longer have a source and are discontinued for that subset. Indications of implementation remain supported by the lecturer's report and the radiology project, but are not sufficient to conclude maintenance. The mapping conclusions are more persistent than the specific program or coefficient examples, but this does not mean that the stable effect or bias has been eliminated.

Competencies, Learning, and Transfer as Different Outcomes

The heterogeneity of the instruments explains why AI literacy research can produce relationships that look similar but have different meanings. Self-assessments reflect perceptions of ability, concept tests test specific knowledge, and task demonstrations can examine the use as well as evaluation of outputs; combining all three without regard to domains makes a change in beliefs appear to be a change in ability. For HR development, needs diagnosis can juxtapose self-assessment with authentic tasks, such as choosing the right use of AI, correcting prompts, checking sources, and rejecting problematic outputs with the same rubric. These rubrics and tasks are evaluation design recommendations, not validated instruments, and differences in sectors and languages still require local validity testing.

Professional patterns of practice, reflection, and projects provide a basis for structuring learning that is close to the task, but do not determine one of the most effective methods.

Practice-oriented programs can increase the opportunity to use skills without guaranteeing transfer after course support ends (Baldwin & Ford, 1988). HRD evaluations can distinguish between immediate assessment, use on new assignments, and repeated implementation after the program, with usage records complemented by accuracy and quality assessments, as high frequency of use can arise with errors or dependencies. Observation time is determined according to the task cycle and organizational risk; these recommendations are not claimed to be the optimal schedule that has been tested.

Conceptual Framework and Research Propositions

The framework in Figure 2 brings together training, competence, transfer, and performance, as well as the direct relationship of competence to performance as an alternative explanation. Organizational support and task suitability are placed as conditions that need to be tested. The framework is a conceptual integration of authors that have not been tested together on the review data, and the dotted arrows mark propositions, not new mediating or moderation estimates.

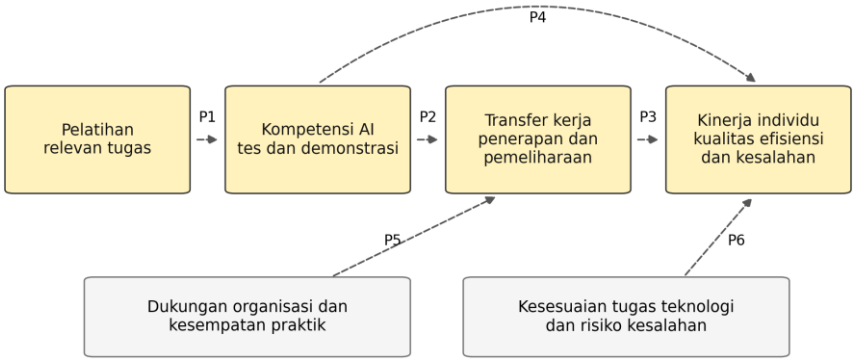


Figure 2. Conceptual framework of AI literacy and work outcomes

Table 9. Propositions for further testing

Code	Propositions	Testing needs
P1	Task-based training is expected to improve AI competencies measured by standardized tests or demonstrations.	Equivalent AI access, active comparators, objective initial and final assessments, dosage and compliance documentation
P2	AI competencies are expected to support the appropriate application to new tasks and their maintenance, when practical opportunities are available.	Competency measured prior to implementation; observation of work tasks, usage records, independent assessors, follow-up
P3	The proper and maintained application of AI competencies is expected to be related to the quality and efficiency of work, without adding to errors.	Measures of real task performance after transfer, comparators, baselines, and errors; not frequency of use alone
P4	The direct relationship of AI competency to performance needs to be tested along the pathway through transfer to distinguish alternative explanations.	Separate competencies from efficacy, temporal sequences, other performance sources, and relevant covariates

Code	Propositions	Testing needs
P5	Organizational support and opportunities to apply competencies are expected to strengthen the competency relationship with transfers.	Size of support and opportunities before transfer; statistical interaction and clear individual or team level
P6	The suitability of AI capabilities with the task is expected to determine the benefits of deployment; risky tasks require human verification.	Variety of task types, error assessment, recorded tool versions, contrast of separate access and training

Testing needs to equalize tool access between groups, measure capabilities prior to deployment, and assess results from other sources when possible. When a tool or task changes, versions and usage conditions need to be noted so that performance changes are not entirely attributed to worker competencies.

Implications of Human Resources Development (RQ5)

The RQ5 answer is an HRD design that links competency diagnosis to application opportunities and assessed task outcomes. Materials can include AI capability limits, tool briefing, output evaluation, and professional responsibilities, with sector-specific examples. Employer support, practice time, access to learning resources, and feedback can be designed as evaluated components. Evaluations should collect perceptions and demonstrations of tasks at the diagnosis stage; materials, doses, attendance, and feedback at the learning stage; task opportunities and barriers to implementation support; observation of new tasks and post-program iterations for transfer; and quality, time, errors, and independent assessments of performance, so that frequency of use is not considered an improvement in outcomes. The review provides a basis for component selection, without ensuring the effectiveness of the combined package.

Limitations and Research Agenda

Web coverage and selective citations may miss reports that are not indexed or difficult to access, and 64 reports are not obtained in sufficient text so that access may affect the corpus. Search engine results do not support claims of a thorough search on the planned database, and the update of two queries is not evidence of literature saturation. The results cover different professions and regions, so recommendations for Indonesian organizations require contextual suitability assessments. The appraisal and AI-assisted extraction have not been verified by two independent humans; 249 additional fields in the initial matrix have not been confirmed, and the extraction of CI, SE, and zero and negative results is not evenly matched. Uncertainty of overlap, denominators, pre- and post-pairs, and some source numbers prevent stronger quantitative claims, while the certainty of formal evidence and unavailable result bias are not statistically assessed.

Subsequent research needs to use task-relevant capability measures, active comparators with equivalent AI access, separate performance appraisal sources, and repeated observations after the program. Preregistration can limit selective reporting, and researchers need to report positive, zero, and negative results, including increased errors. To strengthen this review, the priority is to conduct a database search that has not yet been performed, complete results that affect claims, and conduct human examinations of selection, appraisal, and extraction (Table 10).

Table 10. Evidence gaps and advanced research designs

Gap	Inference implications	Recommended plans
Heterogeneous competency measures	Perception, test, and proxy scores aren't automatically comparable	Validation of constructs and professional tasks; compare abilities with self-assessment
Non-random multiple-choice evaluation	Change can be mixed with confounders or history	Active comparators, equivalent tool access, baseline, and relevant covariates
Maintenance is not yet clear	Immediate results have not proven sustainable transfers	Repeated follow-up on work tasks is different from the precision rubric
Different performance sources and levels	Organizational efficacy or outcomes are not representative of all individual performance	Measures of quality, time, and errors from separate raters
Support and contextual norms	Qualitative themes don't automatically prove moderation	Measure support and opportunities before transfers; test interactions at a clear level
Number mismatches and overlaps	Denominators and effect estimates can be incorrectly combined	Document pairs, sample units, report versions, CI/SE, and sample relationships

Source: thematic synthesis. The design is a recommendation, not an empirical activity that has been carried out by the review.

CONCLUSION

Employee artificial intelligence (AI) literacy was operationalized using diverse measures of competencies, self-reported perceptions, and proxy indicators. Professional training programs incorporated practical exercises and context-specific projects; however, the available evidence was insufficient to identify the most effective training approaches. Although several studies reported the application of AI-related competencies in workplace settings, sustained competency retention and training transfer following program completion remained inadequately established. The relationships between AI literacy and workplace outcomes varied according to the constructs examined, research designs, and organizational conditions. Consequently, AI adoption readiness, employee well-being, and job performance should be treated as distinct outcomes rather than interchangeable indicators.

Human resource development practices could strengthen the relationship between AI literacy and employee performance by integrating competency assessments, targeted professional training, workplace application support, and objective evaluations of task quality and error rates. The proposed conceptual framework and six research propositions provided a foundation for future empirical investigations into the relationships among AI competencies, training transfer, and job performance. Nevertheless, the conclusions were limited to the mapping and interpretation of the available evidence and did not establish universal effects or definitive causal relationships. Comparisons across study subsets indicated that some overarching findings remained consistent, whereas specific outcomes varied depending on the characteristics, methodological approaches, and evidence reported in individual conference proceedings.

FUNDING

This research was carried out independently by the authors, without any financial support, grants, or financing in any form from any institution, organization, or party during the research and manuscript writing process.

CONFLICT OF INTEREST

The authors state that there is no conflict of interest in this study. The author also does not have commercial, financial, or other relationships with certain parties that have the potential

to affect the conduct of the research, interpretation of the results, or presentation of the findings in this manuscript.

AUTHOR CONTRIBUTION

All authors contribute in a balanced and equal manner in the completion of this manuscript. These contributions include the preparation of research ideas and frameworks, data analysis, to the writing and improvement of the manuscript, so that there is no division of dominant roles among the authors.

GENERATIVE ARTIFICIAL INTELLIGENCE USAGE STATEMENT

AI is used to assist with source search, registry organization, selection and appraisal decision proposals, extraction, consistency checks, and text drafting. AI is not treated as a second-person author or assessor. The substantive decisions in this package have not yet received independent human adjudication. The supplemental document contains a reporting checklist, search log, update register, report matrix, appraisal profile, claim traceability, and submission readiness record. The author's identity, human contribution, funding, and conflict of interest need to be disclosed by the author in a separate submission file; their absence is not assumed.

REFERENCES

Alghowinem, S., Bagiati, A., Salazar-Gómez, A. F., & Breazeal, C. (2024). Leading in the AI era: An interactive experiential hands-on learning approach for professionals and leaders. 2024 ASEE Annual Conference & Exposition, Paper 47719. https://doi.org/10.18260/1-2--47719

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). Computers and Education Open, 6, Article 100173. https://doi.org/10.1016/j.caeo.2024.100173

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41(1), 63–105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2024). Explicating AI literacy of employees at digital workplaces. *IEEE Transactions on Engineering Management*, *71*. https://doi.org/10.1109/TEM.2021.3138503

Chandra, W. (2026). Building employee agility in the AI-enabled workplace: Digital leadership, AI literacy, digital HRM practices, and the moderating role of digital mindset. Journal of Economic and Leadership, 3(2), 70–96. https://doi.org/10.70142/jbl.v3i2.71

Che, C. (2026). AI literacy as a meta-skill: A four-domain model for academic management innovation in higher education. Frontiers in Education, 11, Article 1755238. https://doi.org/10.3389/feduc.2026.1755238

Çoban, N., & Beydağ, K. D. (2026). Surgical nurses' artificial intelligence literacy and readiness levels for medical artificial intelligence. BMC Nursing, 25(1), Article 176. https://doi.org/10.1186/s12912-025-04248-6

Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organization Science, 37(2), 403–423. https://doi.org/10.1287/orsc.2025.21838

Dolezel, D., Lambert, E., Watzlaf, V., Morton, M., Lalani, K., Sand, J., & Fenton, S. (2025). Assessing artificial intelligence literacy: What do health information professionals know about AI? Advances in Health Information Science and Practice, 1(2), Article KLVV3078. https://doi.org/10.63116/KLVV3078

Fu, Y., Zhou, Q., Shi, G., Chen, C., Jiang, E., & Ma, J. (2026). Inclusive leadership, psychological safety, AI readiness, and nurses' AI literacy: A longitudinal study. Journal of Nursing Management, 2026, Article 9935514. https://doi.org/10.1155/jonm/9935514

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). Mixed methods appraisal tool (MMAT), version 2018: User guide. McGill University. https://medschool.cuanschutz.edu/docs/librariesprovider94/di-docs/methods-(design)-docx/mmat_2018_criteria-manual_2018-08-01_eng.pdf

Hutabarat, H., Djuningsih, I., Astri, M., Astuti, P., & Irawan, S. (2025). Adaptasi skala literasi kecerdasan artifisial (SAIL4ALL) ke dalam bahasa Indonesia: Studi pada karyawan perusahaan retail di Indonesia. Jurnal Kesehatan dan Pengembangan Terbaru, 6(3), 35–62. https://ejurnals.com/ojs/index.php/jkpt/article/view/2892

Kalff, Y., & Simbeck, K. (2025). Explained, yet misunderstood: How AI literacy shapes HR managers' interpretation of user interfaces in recruiting recommender systems. RecSysHR 2025, CEUR Workshop Proceedings, 4046. https://ceur-ws.org/Vol-4046/RecSysHR2025-paper_3.pdf

Karn, D. K., Mishra, A. K., Shah, B., Dhakal, J. K., & Aithal, P. S. (2026). Generative AI usage, AI literacy, and employee performance in Parsa's industrial and trading enterprises. Poornaprajna International Journal of Emerging Technologies (PIJET), 3(2), 118–131. https://doi.org/10.64818/pijet.3107.8486.0027

Khalil, A. I., Bakheet, J., Atiya, D., Al Jehani, R., Abdullah, R., & Haddad, M. (2025). Cultivating artificial intelligence (AI) competence and shaping attitudes among psychiatric hospital nurses: A quasi-experimental study. Digital Health, 11, Article 20552076251336515. https://doi.org/10.1177/20552076251336515

Kong, S. C., Mak, Z. W. S., Wu, Y., & Yang, Y. (2024). Evaluating the effectiveness of a professional development course on artificial intelligence literacy for administrative staff in higher education. Proceedings of the 32nd International Conference on Computers in Education. https://doi.org/10.58459/icce.2024.4985

Kong, S. C., Yang, Y., & Wu, Y. (2025). Evaluation of a professional development course on artificial intelligence literacy for administrative staff in Hong Kong. The Global Chinese Conference on Computers in Education Proceedings. https://nlt.gcsce.net/ojs/index.php/GCCCE/article/view/587

Kosonen, M., & Saltiola, K. (2026). Bridging the AI competence gap: Mapping skills and training needs in SMEs of technology and creative sectors. Finnish Business Review. https://verkkolehdet.jamk.fi/finnish-business-review/2026/05/19/bridging-the-ai-competence-gap-mapping-skills-and-training-needs-in-smes-of-technology-and-creative-sectors/

Lastovska, A., Lāma, G., Kalkis, H., Roja, Z., & Rusakova, A. (2026). AI as a human complement in the labour market: Comparing employees' AI competence in finance, healthcare, and public services. Human-Intelligent Systems Integration, Advance online publication. https://doi.org/10.1007/s42454-026-00095-1

Lee, J.-H., & Noh, H.-H. (2026). Individual competencies driving innovative work behavior in generative AI–enabled environments: AI literacy, expertise and metacognition. Asia-Pacific Journal of Business Administration, Advance online publication. https://doi.org/10.1108/apjba-10-2025-0781

Li, Y., & Mao, Z. (2026). AI identity threat and creative teaching among college teachers: A job demands-resources mechanism with the moderating roles of AI literacy and organizational support. Sage Open, 16(4), Article 21582440261493838. https://doi.org/10.1177/21582440261493838

Liu, W., Li, Y., Yan, Y., Wang, J., Wang, J., & Zhang, H. (2026). Exploring university faculty's AI well-being: A structural equation model of social supports, AI literacy, and technological self-efficacy. Behavioral Sciences, 16(7), Article 1168. https://doi.org/10.3390/bs16071168

Liu, X., Zhang, L., & Wei, X. (2025). Generative artificial intelligence literacy: Scale development and its effect on job performance. Behavioral Sciences, 15(6), Article 811. https://doi.org/10.3390/bs15060811

Liu, Y., Zhao, X., & Zeng, Q. (2026). AI literacy as a potential mediator between attitude and self-efficacy among PICU nurses: A cross-sectional study. Frontiers in Public Health, 14, Article 1812848. https://doi.org/10.3389/fpubh.2026.1812848

Lo, L. S. (2024). Evaluating AI literacy in academic libraries: A survey study with a focus on U.S. employees. College & Research Libraries, 85(5), 635–668. https://doi.org/10.5860/crl.85.5.635

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

Looi, C.-K., Yang, Y., Velander, J., Milrad, M., & Sun, D. (2026). Preparing educators for the AI-enhanced future: Insights from a teacher professional development for K-12 education in Singapore. Research and Practice in Technology Enhanced Learning, 21, Article 41. https://doi.org/10.58459/rptel.2026.21041

Naveed, M. A., Asghar, M. Z., Talha, Riaz, M., & Zeshan, A. (2026). Symmetric and asymmetric nexus between AI literacy, employee engagement, and work performance among academicians. Discover Computing, 29(1), Article 258. https://doi.org/10.1007/s10791-026-10145-6

Nguyen, K. M., Bui, T. T. M., Pham, H. T. T., Nguyen, L. X., Pham, H. T. X., Le Gia, L., & Nguyen, N. T. (2026). Fostering employees' AI adoption in strategic marketing planning and decision making: A mixed-method study in Vietnam. Cognition, Technology & Work, 28(1), 169–202. https://doi.org/10.1007/s10111-025-00838-1

Novitasari, D., Anshori, M. Y., & Has Shiela, F. J. (2025). The effect of digital leadership on performance through task-technology fit and AI literacy in creative industries. Jurnal Ilmiah Manajemen Kesatuan, 13(6), 5417–5428. https://doi.org/10.37641/jimkes.v13i6.3832

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, Article n71. https://doi.org/10.1136/bmj.n71

Passlack, N., Hammerschmidt, T., & Posegga, O. (2026). With great power comes great responsibility: What shapes AI literacy for responsible interactions of knowledge workers with AI? Information Systems Frontiers, 28(1), 11–46. https://doi.org/10.1007/s10796-025-10648-5

Pinski, M., Hofmann, T., & Benlian, A. (2024). AI literacy for the top management: An upper echelons perspective on corporate AI orientation and implementation ability. Electronic Markets, 34(1), Article 24. https://doi.org/10.1007/s12525-024-00707-1

Rabbani, S., Siddiqui, A., Khan, S. W., & Halim, A. (2026). AI instructional prescriptiveness and adaptive performance: Evaluating the instructional agency paradox through cognitive dependency and perceived autonomy. Journal of Management Practices, Humanities and Social Sciences, 10(3), 392–402. https://doi.org/10.33152/jmphss-10.3.5

Riggs, V. (2025). Impact of AI literacy professional development on teaching: Evaluating changes in perceptions and classroom practices. American Journal of STEM Education, 7, 37–50. https://doi.org/10.32674/rbajg012

Sebok, A., Merchant, M., & Tsai, C.-L. (2025). Changes in perceptions and experiences following a generative AI learning program. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 69(1), 888–893. https://doi.org/10.1177/10711813251363183

Setiana, A. R., Kalsum, E. U., & Jatnika, R. A. (2026). Kesiapan kerja karyawan room division pada era artificial intelligence peran AI literacy dan digital competence. Journal of Management and Social Sciences, 5(1), 75–92. https://doi.org/10.55606/jimas.v5i1.2878

Steinhauser, S., & Heid, P. (2026). Organizational readiness, AI literacy, and the new frontier of R&D: How generative AI shapes innovation capacity. R&D Management, 56(4), 559–578. https://doi.org/10.1111/radm.70038

Tan, W., Nawaz, M., Shu, T., & Ramzan, B. (2026). Collaborative artificial intelligence literacy and employee performance: Task–technology fit and technostress in a moderated mediation model. South African Journal of Business Management, 57(1), Article a5903. https://doi.org/10.4102/sajbm.v57i1.5903

van de Venter, R., Skelton, E., Matthew, J., Woznitza, N., Tarroni, G., Hirani, S. P., Kumar, A., Malik, R., & Malamateniou, C. (2023). Artificial intelligence education for radiographers, an evaluation of a UK postgraduate educational intervention using participatory action research: A pilot study. Insights into Imaging, 14(1), Article 25. https://doi.org/10.1186/s13244-023-01372-2

van Kooten, M. J., Tan, C. O., Hofmeijer, E. I. S., van Ooijen, P. M. A., Noordzij, W., Lamers, M. J., Kwee, T. C., Vliegenthart, R., & Yakar, D. (2024). A framework to integrate artificial intelligence training into radiology residency programs: Preparing the future radiologist. Insights into Imaging, 15(1), Article 15. https://doi.org/10.1186/s13244-023-01595-3

Wiggins, W. F., Caton, M. T., Magudia, K., Glomski, S.-H. A., George, E., Rosenthal, M. H., Gaviola, G. C., & Andriole, K. P. (2020). Preparing radiologists to lead in the era of artificial intelligence: Designing and implementing a focused data science pathway for senior radiology residents. Radiology: Artificial Intelligence, 2(6), e200057. https://doi.org/10.1148/ryai.2020200057

Xia, Q. N., Constantinides, M., Sarkar, A., Brumby, D. P., & Cox, A. L. (2026). "If you're very clever, no one knows you've used it": The social dynamics of developing generative AI literacy in the workplace. Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work, 1–18. https://doi.org/10.1145/3808045.3808050

Xie, Z., Han, J., Kuang, D., Hu, J., Liao, C., & Wang, R. (2026). Artificial intelligence literacy and associations with thriving at work among nurses in Anhui Province, China: A latent profile analysis. BMC Nursing, 25(1), Article 920. https://doi.org/10.1186/s12912-026-05215-5

Xu, S., Zhang, G., & Xu, W. (2026). AI literacy and innovative work behavior among Chinese university teachers: Indirect associations through AI-TPACK and moderation by organizational inertia. Frontiers in Psychology, 17, Article 1828350. https://doi.org/10.3389/fpsyg.2026.1828350

Yang, Z., & Lan, Y. N. (2026). The chain-mediating roles of generative AI literacy and TPACK in the relationship between AI support and teaching practice among university teachers. Frontiers in Psychology, 17, Article 1932185. https://doi.org/10.3389/fpsyg.2026.1932185

Zhang, A., & Lee, M. K. (2025). Knowledge workers' perspectives on AI training for responsible AI use. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3706598.3714100

Zhang, Q., Zhang, J., & Dong, S. (2026). Dual pathways of generative AI use: Role ambiguity and self-efficacy in employee-AI collaboration. Frontiers in Psychology, 17, Article 1793095. https://doi.org/10.3389/fpsyg.2026.1793095

Zhang, Y., Liu, X., Yan, Q., & Na, M. (2025). Empowering workforces in AI-driven environments: Co-skilling, organizational support, and mitigating job insecurity. Frontiers in Psychology, 16, Article 1700129. https://doi.org/10.3389/fpsyg.2025.1700129