



AI-Assisted Educational Interventions in Medical Imaging Education: A Systematic Review of Learning Outcomes, Diagnostic Competencies, and Implementation

Leigh Kirsten A. Ybañez-Lauron*, Reanne Clariss M. Rayco-Caya, Ian P. Laping, Mary Ruassel A. Rejollo, Roberto B. Tagalog Jr

Department of Radiologic and Medical Technology, College of Allied Medical Sciences, Cebu Doctors' University, Mandaue City, Cebu Philippines

DOI: <https://doi.org/10.68050/JAMS.2026.516>

Received: 17 Sept 2026; Accepted: 18 Sept 2026; Published: 26 September 2026

ABSTRACT

Background: Artificial intelligence (AI) is increasingly integrated into medical-imaging education to support image interpretation, personalised learning, feedback, simulation, and competency development. However, evidence regarding its educational effectiveness and implementation remains heterogeneous.

Objective: This systematic review evaluated AI-assisted educational interventions in medical-imaging education, focusing on learning outcomes, diagnostic competencies, and learner-reported and implementation outcomes.

Methods: Electronic databases and supplementary sources were searched for studies evaluating learner-facing AI-assisted educational interventions across medical-imaging disciplines. Eligible controlled, uncontrolled, single-group, and within-participant intervention studies were included. Findings were synthesised narratively because of substantial heterogeneity in learner populations, AI technologies, study designs, interventions, comparators, and outcomes. Objective effectiveness outcomes were considered separately from learner-reported and implementation outcomes.

Results: Seventeen unique studies met the eligibility criteria. Controlled studies generally reported favourable effects of AI-assisted approaches on image interpretation, diagnostic performance, personalised learning, and feedback. Uncontrolled studies also reported improvements in knowledge or diagnostic competency, although causal interpretation was limited by methodological limitations. Learner-reported and implementation findings were generally favourable, particularly for confidence, perceived knowledge, usefulness, and acceptability. Important concerns included hallucinations, reliability, technological overreliance, accountability, data security, and clinical fidelity. Meta-analysis was not undertaken because of substantial heterogeneity.

Conclusion: AI-assisted educational interventions show promise for improving selected learning and diagnostic outcomes in medical-imaging education. However, heterogeneous methods, small samples, uncontrolled designs, short follow-up, and reliance on self-reported outcomes limit certainty. AI should therefore complement, rather than replace, structured educator-led instruction.

Keywords: Artificial intelligence, Medical imaging education, Radiology education, Diagnostic accuracy, Educational interventions, Generative artificial intelligence.

INTRODUCTION

Artificial intelligence (AI)-assisted education refers to the structured use of AI technologies to support teaching, learning, assessment, feedback, and the development of knowledge or professional competencies (Masters et al., 2025). It encompasses interventions based on machine learning, natural-language processing, computer vision, learner modelling, and educational data analytics. These technologies can support persona-



lized feedback, automated assessment, instructional recommendations, simulated practice, adaptive learning pathways, and interactive learning experiences (Chen et al., 2020; Zawacki-Richter et al., 2019). AI-assisted educational interventions therefore include adaptive learning platforms, intelligent tutoring systems, AI-generated feedback, automated assessment systems, conversational agents and large language models (LLMs), virtual simulations, and AI-supported problem- or case-based learning (Matua et al., 2026; Suazo-Galdamés & Chaple-Gil, 2025). Rather than replacing educators or established teaching methods, these interventions may complement lectures, demonstrations, supervised practice, assigned readings, and conventional feedback by providing more immediate, interactive, and individualised learning opportunities (Létourneau et al., 2025).

The educational effects of AI-assisted interventions can be evaluated across several domains. Academic outcomes include examination performance, knowledge gain, course achievement, and short- or long-term retention, whereas competency-based outcomes include image-interpretation accuracy, diagnostic accuracy, anatomical and pathological recognition, clinical reasoning, image-quality evaluation, report writing, procedural decision-making, and simulated or clinical performance (Chau et al., 2022; Matua et al., 2026). Learner-reported and implementation outcomes, including engagement, motivation, confidence, satisfaction, usability, cognitive load, learning efficiency, time on task, self-regulated learning, and acceptability, can provide complementary information about the feasibility and educational experience of an intervention (du Plooy et al., 2024; Lai et al., 2026). However, these outcomes should be distinguished from objectively measured improvements in knowledge or competency. Interpretation of educational effectiveness must also consider study design, baseline comparability, intervention duration, assessment validity, alignment between the intervention and measured outcome, and timing of follow-up assessment (Kulik & Fletcher, 2016).

AI-assisted education has become increasingly relevant because learners differ in prior knowledge, digital competence, learning pace, language proficiency, and access to educational support (Luo et al., 2025). AI technologies may provide repeated practice, immediate feedback, personalised learning pathways, and targeted support for individual knowledge deficiencies (Suazo-Galdamés & Chaple-Gil, 2025). Nevertheless, educational effectiveness is not uniform. A meta-analysis of 45 independent studies reported positive effects of AI-enabled adaptive learning on cognitive, affective, and behavioural outcomes, although effect sizes varied according to learner level, intervention design, and implementation conditions (Wang et al., 2024). Similarly, a review of 69 higher-education studies identified improved academic performance in 41 studies, while 28 reported no significant improvement (du Plooy et al., 2024). Evidence from health-professions education also suggests potential improvements in diagnostic accuracy, knowledge retention, personalised feedback, and learning efficiency, although AI-assisted approaches do not consistently outperform conventional teaching across all outcomes (Matua et al., 2026). These findings indicate that educational effectiveness is influenced not only by the AI technology itself but also by instructional design, educator involvement, learner characteristics, assessment methods, technological infrastructure, and alignment with curricular objectives.

Medical-imaging education represents a particularly relevant setting for AI-assisted learning because learners must integrate knowledge of anatomy, physiology, pathology, imaging physics, radiation protection, image production, patient care, and clinical decision-making (ISRRT, 2004). Beyond foundational knowledge, learners require repeated opportunities to interpret images, recognise normal and abnormal findings, evaluate image quality, formulate differential diagnoses, prepare reports, make procedural decisions, and receive timely and meaningful feedback. AI-assisted interventions may support these activities through interactive image cases, automated or personalised feedback, adaptive exercises, virtual simulation, diagnostic-support demonstrations, synthetic-image training, and progressive assessment (Duong et al., 2019; Keshavarz et al., 2026). The increasing clinical integration of AI into medical imaging further requires future professionals to understand how to use algorithmic outputs critically and recognise potential errors, biases, ethical concerns, and governance requirements (Melford & O'Brien-Melford, 2026; Cayabyab et al., 2026). Medical-imaging curricula have consequently been encouraged to incorporate AI principles, clinical applications, data governance, ethical considerations, and critical appraisal of AI-supported decisions (Crotty et al., 2024).

These educational considerations extend across radiography, radiologic technology, diagnostic imaging, sonography, nuclear medicine, dental imaging, medical-student radiology instruction, and postgraduate radiology training. However, educational needs and implementation conditions differ across undergraduate



students, residents, fellows, and practising professionals. The effectiveness and feasibility of AI-assisted interventions may likewise vary according to imaging discipline, learner experience, educational function of the AI tool, duration of exposure, educator involvement, assessment method, and availability of technological resources. These challenges may be particularly relevant in settings where digital infrastructure and institutional preparedness are uneven. In Philippine higher education, for example, interest in AI-assisted teaching, assessment, and research is increasing, but implementation varies according to digital infrastructure, internet and electricity reliability, faculty preparedness, digital literacy, institutional policies, and learner access to suitable devices and services (Cacho, 2024; Co, 2025; UNESCO, 2025). A mixed-method study involving 451 students from a rural Philippine higher-education institution reported widespread AI-tool use and generally favourable perceptions, alongside concerns regarding inaccurate information, reduced critical thinking, academic dishonesty, and limited awareness of institutional policies (Villarino, 2025). These considerations highlight the importance of evaluating not only whether AI-assisted interventions improve educational outcomes but also the conditions under which they can be implemented appropriately.

Despite rapidly growing interest, evidence on AI-assisted medical-imaging education remains heterogeneous and fragmented. Existing studies have examined different learner populations and educational functions, including adaptive learning, automated and personalised feedback, conversational and generative AI, AI-supported image interpretation, computer-aided diagnosis, virtual or AI-assisted simulation, and synthetic-image training (Biswas, 2023; Lewis et al., 2024; Salastekar et al., 2023). Studies also differ substantially in design, comparator, intervention duration, assessment instrument, follow-up period, and definition of educational effectiveness (Doherty et al., 2024; Garin et al., 2023; Lai et al., 2026). Importantly, much of the literature on AI in medical-imaging education has focused on awareness, attitudes, readiness, acceptance, or curriculum preferences rather than evaluating a defined learner-facing intervention and its effects on measurable educational outcomes (Arif, 2024; Li et al., 2025). Previous reviews have also frequently combined multiple healthcare disciplines or concentrated primarily on radiology trainees, limiting conclusions across the broader spectrum of medical-imaging learners and professionals (Keshavarz et al., 2026; Matua et al., 2026; Suazo-Galdamés & Chaple-Gil, 2025).

A further limitation of the existing literature is that objective evidence of educational effectiveness is often considered alongside self-reported confidence, satisfaction, readiness, or perceived usefulness without sufficient distinction between these outcome domains. Similarly, findings from controlled trials may be considered together with uncontrolled pre-test/post-test studies despite important differences in their ability to support causal inference. A systematic synthesis that distinguishes objective learning and diagnostic competency from learner-reported and implementation outcomes, while also separating controlled from uncontrolled evidence, is therefore needed to provide a clearer assessment of the educational value of AI across medical-imaging disciplines.

Accordingly, this systematic review evaluated AI-assisted educational interventions in medical-imaging education and synthesised their effects on learning outcomes, diagnostic competencies, and implementation outcomes. Specifically, the review examined objective outcomes including knowledge and examination performance, image interpretation and diagnostic accuracy, reporting performance, clinical reasoning, procedural competency, and knowledge or skill retention. Learner-reported and implementation outcomes, including engagement, confidence, motivation, satisfaction, usability, learning efficiency, and acceptability, were examined separately. The review further characterised interventions according to learner population, imaging discipline, AI educational function, study design, comparator, and educational setting, while distinguishing controlled evidence from uncontrolled intervention studies. Methodological quality was assessed using design-appropriate JBI critical appraisal tools and considered in interpreting the evidence, with the aim of identifying current evidence gaps and informing priorities for future research and the responsible integration of AI-assisted approaches into medical-imaging education.

METHODOLOGY

Study Design

This systematic review evaluated AI-assisted educational interventions used in medical-imaging education and examined their effects on learning outcomes, diagnostic competencies, and implementation outcomes. The review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). The review methods were defined before study selection, and any methodological decisions made during the review process were documented to promote transparency and reproducibility.

Eligibility Criteria

Eligibility criteria were defined using the Population, Intervention, Comparator, Outcomes, and Study Design (PICOS) framework. Eligible participants included students enrolled in radiologic technology, radiography, medical imaging, diagnostic imaging, sonography, ultrasound, nuclear medicine, or related programmes. Medical and dental students receiving formal instruction in radiology, radiographic interpretation, or other medical-imaging content were also eligible, as were radiology residents, fellows, trainees, and practising imaging professionals participating in a defined educational programme. Studies involving mixed healthcare learner groups were considered when the intervention primarily concerned medical imaging or when findings relating to eligible participants could be extracted separately.

For the purposes of this review, an AI-assisted educational intervention was operationally defined as an educational activity in which an artificial-intelligence-based system was directly incorporated into learner instruction, practice, assessment, simulation, educationally integrated diagnostic decision support, or feedback. AI technologies used solely for clinical diagnosis, image processing, workflow optimisation, or research without an identifiable learner-facing educational activity were not considered educational interventions. Diagnostic-support technologies were eligible only when they were deliberately incorporated into a defined teaching or training activity and educational or competency-related outcomes were evaluated. Medical-imaging education was defined as formal or structured education or training involving the acquisition, interpretation, evaluation, reporting, or application of knowledge and competencies related to radiology, radiography, radiologic technology, diagnostic imaging, ultrasound/sonography, nuclear medicine, dental imaging, or other eligible imaging-based disciplines.

Eligible interventions included learner-facing AI-assisted tools used to teach, practise, assess, or provide feedback on medical-imaging knowledge and competencies. These included adaptive learning platforms, intelligent tutoring systems, machine-learning or deep-learning applications used for instruction, AI-generated personalised feedback, AI-supported image-interpretation or report-writing training, automated assessment systems, AI-assisted simulation, conversational agents, large language models (LLMs), generative-AI applications, and AI-supported problem- or case-based learning. To be eligible, the intervention had to involve an identifiable educational activity delivered to learners. Studies concerned solely with awareness, attitudes, readiness, perceptions, acceptance, curriculum preferences, or the clinical performance of an AI system without a learner-facing educational component were excluded.

The presence of a comparison group was not mandatory. Where available, eligible comparators included traditional lectures, instructor-led teaching, standard feedback, conventional simulation, non-AI digital learning, self-directed study, usual educational practice, or another educational intervention. Randomised controlled trials, non-randomised controlled studies, quasi-experimental studies, controlled before-and-after studies, within-participant comparative studies, and single-group pre-test/post-test studies were eligible. Studies evaluating a delivered AI-assisted educational intervention primarily through learner-reported, usability, or implementation outcomes were retained for the descriptive implementation synthesis but were not used as evidence of objective educational effectiveness. Reviews, protocols, editorials, commentaries, case reports, descriptive surveys without an evaluated educational intervention, qualitative-only studies, and correlational studies without delivery and evaluation of an educational intervention were excluded. Conference



abstracts or proceedings were considered only when sufficient original data were available to determine eligibility and extract the outcomes required for the review.

Outcome Measures

The primary outcomes were objective measures of learning and competency. These included knowledge or examination scores, image-interpretation accuracy, diagnostic accuracy, anatomical or pathological recognition, clinical reasoning, image-quality evaluation, report-writing performance, procedural decision-making, technical competency, and knowledge or skill retention.

Secondary outcomes included engagement, motivation, confidence, satisfaction, usability, cognitive load, learning efficiency, time on task, self-regulated learning, perceived educational usefulness, and intervention acceptability. Implementation-related outcomes, including barriers, facilitators, concerns regarding AI reliability, technological limitations, and conditions influencing adoption, were also considered where reported. Objective and subjective outcomes were synthesised separately. Studies reporting only learner perceptions, self-assessed knowledge or competency, usability, satisfaction, confidence, or other implementation outcomes were included in the descriptive learner-reported and implementation synthesis but were not interpreted as evidence of objectively improved learning or diagnostic competency.

Information Sources and Search Strategy

PubMed/MEDLINE, Scopus, Web of Science Core Collection, CINAHL, and the Cochrane Library were searched from database inception through 6 September 2026. Supplementary searches were conducted through Google Scholar and other sources to identify additional potentially eligible studies. Only English-language publications were considered. The search strategy was designed to maximise sensitivity and combined three principal concepts: artificial intelligence, education or training, and medical imaging. Learner and outcome terms were not required in the principal search because these concepts may not consistently appear in article titles, abstracts, or indexing terms. Controlled vocabulary and free-text terms were adapted to the requirements of each database.

AI-related search terms included artificial intelligence, machine learning, deep learning, generative AI, large language models, ChatGPT, chatbots, conversational agents, intelligent tutoring systems, adaptive learning, and automated or personalised feedback. Educational terms included education, teaching, learning, training, tutoring, curriculum, feedback, and simulation. Medical-imaging terms included radiology, radiography, radiologic technology, medical imaging, diagnostic imaging, image interpretation, image reporting, computed tomography, magnetic resonance imaging, ultrasound, sonography, mammography, nuclear medicine, and dental imaging. Search terms were combined using appropriate Boolean operators and adapted to the syntax and indexing system of each database.

Google Scholar was used as a supplementary search source, with the first 200 records ranked by relevance screened for eligibility, and supplementary searching was undertaken to identify potentially eligible studies that may have been recently published or inconsistently indexed. Reference lists of included studies and relevant reviews were also examined, and forward citation searching was undertaken where appropriate. Targeted supplementary searches combining generative-AI terms, image-interpretation terms, and medical-imaging education terms were additionally performed. The database and supplementary searches identified a total of 328 records before completion of the study-selection process.

Study Selection

All retrieved records were compiled in a reference-management programme, and duplicate records were removed electronically and verified manually. Before formal screening, the eligibility criteria were reviewed and applied consistently to ensure that studies were assessed according to the predefined inclusion and exclusion criteria. Two reviewers independently screened the titles and abstracts of retrieved records. Potentially eligible reports were subsequently retrieved and assessed in full text. Disagreements regarding



eligibility were resolved through discussion and, when consensus could not be reached, consultation with a third reviewer. Reasons for exclusion at the full-text stage were documented.

Multiple publications arising from the same underlying study or participant sample were linked and treated as a single study to avoid double counting. When both preprint and peer-reviewed versions of the same study were identified, the peer-reviewed publication was treated as the primary report where appropriate. The search and screening process began with 328 identified records, from which 17 unique studies ultimately met the predefined eligibility criteria and were included in the systematic review. The complete identification, screening, eligibility, and inclusion process is presented in the PRISMA 2020 flow diagram (Figure 1).

Data Extraction

A standardised data-extraction form was developed and used to ensure consistent collection of relevant study information. Two reviewers independently extracted or verified data on authorship, publication year, country, educational setting, study design, sample size, participant characteristics, learner type, imaging discipline, AI technology, educational function, intervention content, duration and frequency, educator involvement, comparator, assessment instrument, follow-up period, and reported outcomes.

For objective outcomes, numerical findings were extracted at relevant assessment points, including baseline and post-intervention values, measures of variance, effect estimates, diagnostic-performance measures, and statistical significance where available. Measures such as accuracy, sensitivity, area under the receiver operating characteristic curve, kappa statistics, examination or knowledge scores, and other competency measures were recorded when reported by the original studies.

Information on learner-reported and implementation outcomes was also extracted, including confidence, satisfaction, usability, acceptability, perceived educational usefulness, engagement, implementation barriers and facilitators, and adverse or unintended educational effects. Funding sources, conflicts of interest, and reported study limitations were additionally recorded. When multiple reports described the same study or participant sample, they were treated as a single study, and complementary information was extracted across reports. The most complete peer-reviewed publication was used as the primary source to minimise duplication of study populations and outcomes.

Methodological Quality Assessment

The methodological quality of the included studies was assessed using design-appropriate Joanna Briggs Institute (JBI) critical appraisal tools. Randomized controlled trials were assessed using the JBI Critical Appraisal Checklist for Randomized Controlled Trials, while non-randomized, quasi-experimental, controlled before-and-after, and single-group pre-test/post-test intervention studies were assessed using the JBI Critical Appraisal Checklist for Quasi-Experimental Studies. For studies employing other designs, the corresponding JBI design-specific critical appraisal tool was used where applicable. Item-level judgments were recorded as Yes, No, Unclear, or Not Applicable, as appropriate. Checklist responses were not converted into an arbitrary numerical quality score. Methodological limitations identified through the JBI appraisal were considered when interpreting individual studies and the overall body of evidence. Two reviewers independently conducted the critical appraisal, and disagreements were resolved through discussion and, where necessary, consultation with a third reviewer.

Data Synthesis

The characteristics and findings of the 17 included studies were summarised in tables and synthesised narratively because of substantial heterogeneity in learner populations, AI technologies, medical-imaging disciplines, educational functions, study designs, comparators, intervention durations, and outcome measures. The synthesis was structured according to study design and outcome domain to distinguish evidence of objective educational effectiveness from learner-reported and implementation findings. Controlled studies evaluating objective learning or diagnostic performance were synthesised separately from uncontrolled, single-group, and within-participant studies reporting objective outcomes. A third descriptive synthesis included



studies primarily evaluating learner perceptions, confidence, satisfaction, usability, acceptability, perceived educational usefulness, and other implementation-related outcomes.

Objective effectiveness outcomes included knowledge and examination performance, image interpretation, diagnostic accuracy, diagnostic agreement, reporting performance, clinical reasoning, procedural competency, and retention. Learner-reported and implementation outcomes included confidence, satisfaction, usability, acceptability, perceived usefulness, motivation, and concerns or barriers associated with the use of AI-assisted education. Where a study reported both objective and subjective outcomes, findings were incorporated into the relevant components of the synthesis without counting the study more than once. The direction, magnitude, and consistency of findings were considered alongside study design, methodological limitations, learner population, imaging discipline, AI educational function, outcome type, and educational context. Narrative synthesis was conducted in accordance with the Synthesis Without Meta-analysis (SWiM) reporting guideline (Campbell et al., 2020).

Meta-analysis was not performed because the included studies were insufficiently comparable in learner populations, AI-assisted educational interventions, imaging disciplines, study designs, comparators, intervention characteristics, and outcome measures to support meaningful statistical pooling. Individual study findings and reported effect estimates were therefore presented and interpreted narratively. Findings from controlled studies were given greater interpretive weight when assessing educational effectiveness, whereas findings from uncontrolled and within-participant studies were interpreted more cautiously because of their greater susceptibility to confounding, testing effects, and other methodological limitations.

A formal Grading of Recommendations Assessment, Development and Evaluation (GRADE) assessment was not performed because sufficiently comparable bodies of evidence could not be constructed for the principal effectiveness outcomes. Instead, the strength of the evidence was interpreted qualitatively with consideration of study design, methodological limitations, sample size, consistency of findings, outcome type, and the distinction between controlled and uncontrolled evidence. Learner-reported, implementation, and contextual findings were synthesised descriptively to identify recurring findings relating to usability, acceptability, perceived educational usefulness, barriers, facilitators, technological concerns, and conditions influencing educational adoption.

Ethical Considerations

Ethics approval was not required because this systematic review used data from previously published studies and did not involve direct contact with human participants or access to identifiable personal information. Academic integrity was maintained through accurate citation, transparent reporting of study-selection and synthesis procedures, independent screening and appraisal, and consideration of funding sources and conflicts of interest.

RESULTS

Study Selection

The systematic search identified 328 records. Before screening, 68 records were removed, leaving 260 records for title and abstract screening. Of these, 202 were excluded. Fifty-eight reports were sought for retrieval, of which 12 could not be retrieved. Consequently, 46 full-text reports were assessed for eligibility. Twenty-nine full-text reports were excluded for specified reasons, leaving 17 unique studies that met the eligibility criteria and were included in the systematic review (Figure 1).

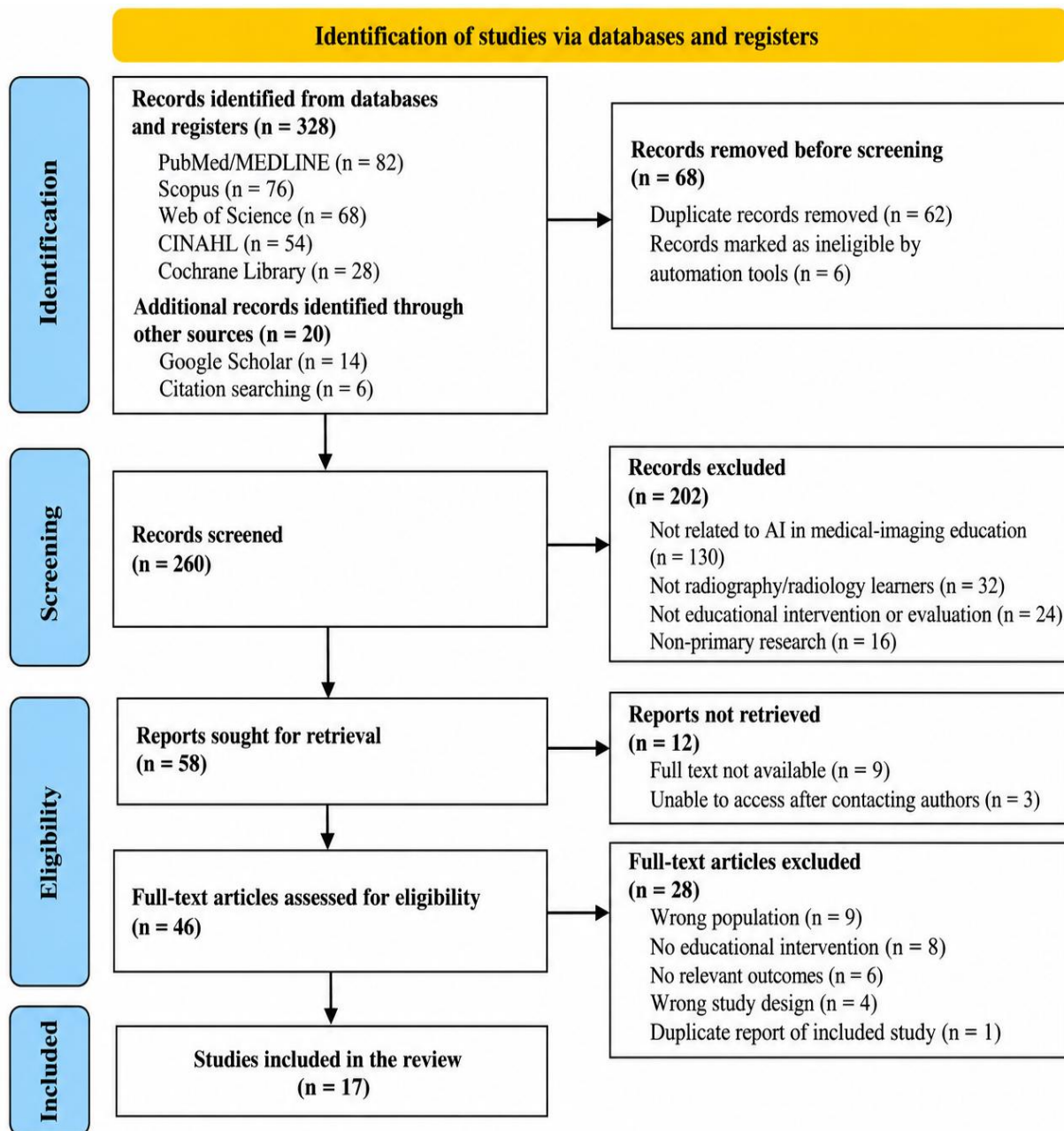


Figure 1. PRISMA 2020 flow diagram of the identification, screening, eligibility assessment, and inclusion of studies in the systematic review.

Study Characteristics

The 17 included studies represented diverse educational settings, learner populations, AI technologies, and medical-imaging disciplines. Participants included medical and dental students, radiography and medical-imaging students, radiology residents and trainees, trainee orthoptists, and practising imaging professionals. The interventions included AI-assisted image interpretation and computer-aided diagnosis, LLM-generated feedback, personalised learning materials, synthetic-image training, structured AI curricula, and generative-AI simulation. Because of substantial heterogeneity in study designs, interventions, and outcomes, the findings were synthesised as controlled studies of objective learning and diagnostic performance, uncontrolled studies of objective learning and diagnostic competency, and studies primarily reporting learner-reported or implementation outcomes.

The principal characteristics, AI-assisted educational functions, outcome measures, and key findings of the 17 included studies are summarized in Table 1.



Table 1. Characteristics of the 17 studies included in the systematic review

Study	Country/setting	Participants	Design	Imaging area	AI intervention / educational function	Comparator	Main outcomes	Key finding
Laupichler et al. (2022)	Germany	24 medical students	Single-group pre/post	Medical imaging	Flipped-classroom AI education with image interpretation and AI-generated diagnoses	None	Self-perceived AI readiness	AI readiness improved; overall comparative self-assessment gain 55.6%.
Hernández-Rodríguez et al. (2023)	Spain	26 radiology professionals (15 specialists, 11 residents)	Educational platform evaluation	Thoracic CT; mammography	JORCAD CNN-based CAD with annotated cases	None	Usability; educational usefulness	Participants generally considered the platform useful for radiology training; objective pre/post competency was not assessed.
Cheng et al. (2020)	Taiwan	30 fifth-year medical students	Controlled educational study*	Pelvic radiography	HipGuide deep-learning assistance for hip-fracture interpretation	Conventional learning	Diagnostic accuracy	Accuracy improvement was greater with AI-assisted learning than conventional learning.
Wang et al. (2024)	China	84 clinical medicine students	Historical-control/non-randomized comparative study	CT; MRI	AI-assisted volume-data reconstruction for diagnostic instruction	Conventional film/multimedia teaching	Practical skills; CT/MRI diagnosis; self-efficacy; self-directed learning	AI group achieved higher practical and diagnostic scores and higher self-



								efficacy/self-directed-learning measures.
Lyu et al. (2024)	China	18 imaging students/junior residents	Controlled longitudinal training study	Pulmonary CT	AI-assisted pulmonary-nodule training	Conventional case-based training	Diagnostic agreement; nodule assessment	Diagnostic agreement improved, particularly among AI-trained imaging students.
Tabuchi et al. (2024)**	Japan	161 trainee orthoptists	Single-group pre/post	Retinal imaging	Synthetic AI-generated ultra-widefield images	None	Image-interpretation accuracy	Accuracy improved for trained ultra-widefield and unseen standard-field images.
Lee et al. (2024)	South Korea	26 inexperienced readers	Within-participant sequential study	Thyroid ultrasound	Digital self-learning followed by AI-CAD assistance	Pre-training performance	Diagnostic performance	AUC improved after self-learning and improved further with AI-CAD assistance.
Finkelstein et al. (2024)	USA	42 radiology residents	Prospective single-group pilot	Radiology/AI	RSNA Imaging AI Foundational Certificate curriculum	None	Objective AI knowledge/skills; familiarity	Objective scores increased from 37% to 73%.
van Kooten et al. (2024)	Netherlands	17 pre-course; 12 post-course residents	Single-group pre/post	Radiology/AI	Three-day AI curriculum	None	Self-rated knowledge/confidence; applicability	Confidence increased from 3.25 to 6.50; 75% considered curriculum applicable.



Kim et al. (2025)	Germany	9 radiology residents	Within-participant comparative study	Brain MRI	LLM search plus structured 10-minute tutorial	Internet search; LLM without tutorial	Diagnostic performance; confidence; time; hallucinations	Top-three diagnostic accuracy reached 62.5% after tutorial versus 44.8% with untutored LLM use and 32.2% with internet search.
Gokkurt Yilmaz et al. (2025)	Türkiye	110 fifth-year dental students	Randomized controlled trial	Dental radiography	ChatGPT-4o personalized study guides	Conventional correct/incorrect feedback	Diagnostic performance; feedback perceptions	Improvement was greater with personalized AI feedback (3.6 ± 1.0 vs 1.3 ± 1.2 ; $p < .001$).
Chen et al. (2025)	China; five hospitals	NR	Multicentre randomized controlled trial	Thyroid ultrasound	AI-assisted diagnostic training	Conventional training; specialty training	Diagnostic accuracy; confidence; emotions	AI and specialty training outperformed conventional training; AI and specialty groups did not significantly differ.
Donkor et al. (2025)	Ghana	144 medical-imaging students	Single-group pre/post	Medical-imaging AI	One-week classroom AI course	None	Objective AI knowledge	Knowledge improved across AI fundamentals, applications, development, and ethics.



Stogiannos et al. (2025)	UK	68 pre; 31 post radiography students	Pre/post survey evaluation	Diagnostic radiography/AI	1.5-hour AI seminar	None	Self-rated knowledge, skills, confidence	Post-seminar ratings were significantly higher, although post-test participation was substantially lower.
Bala et al. (2025)	USA	14 radiology residents	Pilot implementation/reader evaluation	Radiology reporting	GPT-4-generated feedback on preliminary reports	Conventional feedback context	Error detection; satisfaction; perceived usefulness	Model sensitivity was 79.2%; 71.4% favored combined LLM and conventional feedback.
Zhong & Chow (2025)	China	122 students; 155 recent graduates	Pilot educational implementation evaluation	Radiography	Generative-AI imaging simulation	None	Perceptions; realism; educational usefulness	Participants recognized educational potential, particularly for foundational training, while identifying limitations in clinical realism.
Verdone et al. (2026)	USA	8 readers (4 residents, 4 attendings); 100 report pairs	Reader/educational evaluation	Breast imaging	GPT-4o automated feedback on resident reports	Attending consensus	Agreement; feedback helpfulness	GPT agreement with attending consensus ranged 78.3%–



								90.5%; feedback was rated helpful in 83.0%–92.0% of cases.
Chapellier et al. (2026)	Switzerland/Italy/Belgium	13 radiology trainees	Multicentre comparative educational evaluation	Radiology	LLM-generated differential diagnoses	Standard casebook; clinical experience	Perceived usefulness; clarity; trust	Casebook and LLM usefulness rankings were similar; junior trainees showed greater preference for LLM teaching.

Controlled Studies of Objective Learning and Diagnostic Performance

Five studies provided controlled evidence of objective learning or diagnostic performance. Cheng et al. (2020) compared 30 fifth-year medical students receiving conventional or AI-assisted learning using HipGuide, a deep-learning system highlighting suspected hip fractures on pelvic radiographs. Baseline accuracy was comparable between groups. Accuracy in the conventional group increased nonsignificantly from $75.86 \pm 11.36\%$ to $78.66 \pm 14.53\%$ ($p = .264$), whereas the AI-assisted group improved from $75.73 \pm 10.58\%$ at baseline to $88.87 \pm 5.51\%$ during AI-supported testing ($p < .01$) and maintained an accuracy of $84.93 \pm 14.53\%$ during subsequent unaided testing ($p < .01$). The mean accuracy gain was significantly greater in the AI group than in the conventional group (9.2 ± 6.9 vs. 2.8 ± 9.3 percentage points, $p = .042$), with lower-performing students demonstrating the greatest benefit ($r = -.83$, $p < .01$).

Gokkurt Yilmaz et al. (2025) randomised 110 fifth-year dental students to ChatGPT-4o-generated personalised study guides targeting individual learning gaps or conventional correct/incorrect feedback during radiographic-diagnosis training. At one-month reassessment, improvement was significantly greater in the AI-personalised feedback group than in the control group (3.6 ± 1.0 vs. 1.3 ± 1.2 points, $p < .001$), and final diagnostic scores were also significantly higher in the intervention group ($p < .001$). Students receiving AI-personalised feedback additionally rated it as more understandable, beneficial, and motivating than conventional feedback.

Wang et al. (2024) evaluated AI-assisted image-diagnosis instruction among 84 clinical medicine students using a non-randomised historical-control design. Students receiving AI-assisted volume data reconstruction achieved higher total practical-skill scores than those receiving conventional film-based and multimedia instruction (84.51 ± 3.81 vs. 80.67 ± 5.43), including higher CT diagnostic scores (49.93 ± 3.59 vs. 46.60 ± 4.89) and MRI diagnostic scores (17.41 ± 1.00 vs. 16.93 ± 1.14). Academic self-efficacy (82.17 ± 4.67 vs. 78.93 ± 6.29) and self-directed learning scores (235.56 ± 13.50 vs. 226.35 ± 13.90) were also higher in the AI-assisted group.

Lyu et al. (2024) compared seven weeks of pulmonary-nodule training among medical-imaging students receiving conventional case-based instruction, medical-imaging students receiving AI-assisted training, and junior radiology residents receiving AI-assisted training. Diagnostic agreement improved across all groups, with the largest increase among AI-assisted medical-imaging students, whose mean kappa increased from 0.417 to 0.757. This compared with an increase from 0.478 to 0.658 among traditionally trained students and from 0.638 to 0.791 among AI-assisted residents. AI-assisted groups also demonstrated better detection of several nodule categories and fewer false-positive findings.

Chen et al. (2025) conducted an 18-month multicentre randomised controlled trial across five residency-training hospitals comparing conventional training, conventional plus specialty training, and conventional plus AI-assisted training for thyroid-nodule ultrasound diagnosis. Diagnostic accuracy increased progressively in all three groups, but specialty and AI-assisted training resulted in significantly greater accuracy than conventional training alone ($p < .001$). No significant difference was observed between specialty and AI-assisted training ($p = .64$). Diagnostic confidence also increased during training, while negative activating emotions decreased and positive activating emotions increased. The total sample size and allocation across study groups were not reported in the available information.

Uncontrolled Studies of Objective Learning and Diagnostic Competency

Five studies reported objective learning or diagnostic outcomes using uncontrolled, single-group, or within-participant designs. Donkor et al. (2025) evaluated a one-week classroom intervention among 144 medical-imaging students with no previous formal training in medical-imaging AI. Significant improvements in knowledge and understanding were observed across AI fundamentals, clinical applications, AI-system development, and ethics (all $p < .001$). Year of study significantly predicted increased post-test understanding ($p = .015$).



Finkelstein et al. (2024) evaluated 42 radiology residents completing the four-month Radiological Society of North America Imaging AI Foundational Certificate. Mean objective AI knowledge and skills scores increased from 37% before the course to 73% afterward ($p < .001$). All 42 residents completed the programme; 74% believed that it improved their familiarity with AI in radiology, and 57% expressed interest in further AI certification.

Kim et al. (2025) evaluated nine radiology residents interpreting challenging brain MRI cases under three conditions: conventional internet searching, use of an LLM-based search engine without specific training, and use of the same tool after a structured 10-minute tutorial. The correct diagnosis appeared among the three leading differential diagnoses in 62.5% of cases after the tutorial, compared with 44.8% before the tutorial and 32.2% using conventional internet searches. The tutorial significantly improved binary diagnostic scores ($p = .042$), numerical diagnostic scores ($p = .016$), and residents' confidence ($p = .006$), without a meaningful change in interpretation time. Hallucinations occurred in 5.1% of LLM queries.

Lee et al. (2024) evaluated 26 inexperienced thyroid-ultrasound readers who completed digital self-learning using 3,000 annotated benign and malignant nodules followed by AI-CAD–assisted interpretation. Across all readers, the area under the curve increased from 0.679 before training to 0.713 after self-learning and 0.758 with AI-CAD assistance ($p < .05$). Among radiology residents, self-learning increased AUC from 0.700 to 0.743 ($p = .016$) and accuracy from 69.9% to 74.2% ($p = .013$). Subsequent AI-CAD assistance further increased residents' accuracy to 77.1%.

Tabuchi et al. (2024) assessed 161 trainee orthoptists completing a web-based retinal-image interpretation course containing 600 synthetic ultra-widefield images generated using a fine-tuned Stable Diffusion model. Accuracy for ultra-widefield images increased from 43.6% before training to 74.1% afterward ($p < .0001$). Accuracy on previously unseen standard-field images similarly increased from 42.7% to 68.7% ($p < .0001$), demonstrating improvement on an image format not used during the training intervention.

Knowledge Retention and Transfer

Evidence regarding retention and transfer of learning was limited. Most included studies assessed outcomes immediately or shortly after the educational intervention, restricting conclusions regarding durability of learning. Gokkurt Yilmaz et al. (2025) included a one-month reassessment and reported sustained improvement in radiographic diagnostic performance among learners receiving AI-personalised feedback. Cheng et al. (2020) demonstrated improved performance during subsequent unaided testing, providing evidence that improvement was not restricted to periods of active AI assistance. Tabuchi et al. (2024) additionally reported improved interpretation of previously unseen standard-field retinal images after training with synthetic ultra-widefield images, suggesting potential transfer beyond the training image format. Nevertheless, evidence regarding longer-term retention, transfer to new clinical contexts, and independent professional performance remained insufficient.

Learner-Reported and Implementation Outcomes

Eight studies primarily contributed evidence relating to AI readiness, confidence, usability, acceptability, perceived educational usefulness, or AI-generated feedback rather than objective changes in learning or diagnostic competency. Laupichler et al. (2022) evaluated a flipped-classroom course among 24 medical students combining online self-study with interactive sessions in which learners interpreted medical images and compared their decisions with AI-generated diagnoses. Self-perceived AI readiness improved significantly across cognition, ability, vision, and ethics, with an overall comparative self-assessment gain of 55.6%.

Stogiannos et al. (2025) evaluated a 1.5-hour AI seminar among final-year diagnostic-radiography students. Mean self-rated theoretical knowledge increased from 3.85 to 6.57, practical skills from 3.44 to 5.39, and confidence in AI from 3.43 to 5.47, with all changes reported as statistically significant. Students nevertheless expressed continuing concerns regarding AI reliability, accountability, data security, patient confidentiality, and technological overreliance. Van Kooten et al. (2024) similarly found that a three-day radiology-residency AI curriculum increased mean self-rated confidence in AI knowledge and understanding from 3.25 ± 1.48 to



6.50 ± 0.90 ($p = .002$), while 75% reported that the curriculum addressed subjects applicable to their radiology work.

Bala et al. (2025) evaluated GPT-4-generated educational feedback based on discrepancies between residents' preliminary reports and attending radiologists' final reports. GPT-4 identified 24 missed diagnoses in the final evaluation, with reported accuracy and sensitivity of 79.2%. Residents reported mean satisfaction and perceived-accuracy scores of 3.50 and 3.64 out of 5, respectively, and 71.4% preferred combining LLM-generated feedback with conventional educator feedback.

Verdone et al. (2026) evaluated a GPT-4o system designed to identify breast-imaging reporting errors and generate automated educational feedback. Agreement with attending consensus was 90.5% for omission or addition of key findings, 78.3% for incorrect or omitted technical descriptors, and 90.4% for inconsistencies between findings and final assessments. AI-generated feedback was considered helpful in 89.8%, 83.0%, and 92.0% of cases across the respective error categories.

Chapellier et al. (2026) evaluated LLM-generated differential diagnoses among 13 radiology trainees across three European academic centres. A standard radiology casebook received the highest mean educational ranking (1.7 ± 0.2), followed closely by LLM-generated teaching (1.8 ± 0.3), with no significant difference between them ($p = .438$). Junior trainees were more likely than advanced trainees to rank LLM teaching first (66.6% vs. 22.1%, $p = .037$). LLM-generated material also received high ratings for clarity (4.4 ± 0.3), trust (4.3 ± 0.3), differential usefulness (4.3 ± 0.4), and diagnostic usefulness (4.2 ± 0.4).

Hernández-Rodríguez et al. (2023) evaluated JORCAD, an educational platform combining convolutional-neural-network-based computer-aided detection with annotated thoracic CT and mammography cases, among 15 specialist radiologists and 11 radiology residents. Participants generally considered the platform useful for training future radiology specialists and valued its combination of annotated cases and AI-generated diagnostic support. The study primarily provided evidence regarding usability and educational acceptability rather than objective changes in knowledge or diagnostic performance.

Finally, Zhong and Chow (2025) evaluated an in-house generative-AI imaging simulator incorporating DALL·E 3 among radiography students, recent graduates, and lecturers. Fifty-eight percent of students and 48% of graduates believed that AI could enhance case-based clinical teaching, although 80% of both groups considered that the simulator required further improvement. Students were more receptive than graduates to its potential substitution for conventional practical training, whereas graduates expressed greater concerns regarding image realism and clinical fidelity. The simulator was considered more appropriate for foundational anatomy and introductory imaging education than for complex diagnostic training.

Methodological Quality Assessment

Methodological quality varied substantially across the 17 included studies. The randomized controlled trials generally provided stronger designs than the uncontrolled and implementation-focused studies, although incomplete reporting of allocation procedures, blinding, sample size, and missing-data handling limited full appraisal in some cases. Cheng et al. (2020) and Gokkurt Yilmaz et al. (2025) had relatively stronger comparative designs but retained methodological concerns related to sample size and incomplete reporting of some trial procedures. Chen et al. (2025) could not be fully appraised because several relevant methodological details were unavailable in the extracted information.

The non-randomized comparative studies were more susceptible to confounding and selection-related limitations. Wang et al. (2024) used historical academic cohorts without random allocation, creating the possibility of temporal and cohort differences influencing the findings. Lyu et al. (2024) included only six participants per group and compared learners with different baseline levels of clinical experience, limiting precision and generalizability.

Among uncontrolled and within-participant studies, common methodological limitations included the absence of comparator groups, small sample sizes, repeated assessment using identical or similar cases, short follow-up



periods, and potential testing or ordering effects. Donkor et al. (2025) used identical pre- and post-test questions, whereas Lee et al. (2024) reassessed the same cases across sequential study phases. Kim et al. (2025) involved only nine residents, further limiting precision and external validity. Although Tabuchi et al. (2024) also used a single-group design, assessment using previously unseen standard-field images provided some evidence of transfer beyond the specific training images.

Implementation-focused studies were generally limited in their ability to establish educational effectiveness because outcomes were predominantly self-reported measures of confidence, satisfaction, usability, acceptability, or perceived usefulness. Several studies also had very small samples or substantial differences between pre- and post-intervention participation. Stogiannos et al. (2025), for example, decreased from 68 participants at baseline to 31 after the intervention, while Van Kooten et al. (2024) decreased from 17 to 12 participants.

Overall, the evidence base was methodologically heterogeneous. Controlled studies generally provided stronger support for educational effectiveness, whereas findings from uncontrolled, within-participant, and implementation-focused studies require more cautious interpretation because of greater susceptibility to confounding, testing effects, attrition, and self-report bias.

Table 2. Methodological Quality and Critical Appraisal of the 17 Included Studies

Table 2A. JBI appraisal of randomized controlled trials

Table 2A. JBI appraisal of randomized controlled trials													
Study	Randomization	Allocation concealment	Baseline similarity	Participant blinding	Deliverer blinding	Assessor blinding	Groups otherwise treated similarly	Complete follow-up	Analyzed as allocated	Same outcome measurement	Reliable measurement	Appropriate statistics	Appropriate trial design
Gokkurt Yilmaz et al. (2025)	Y	U	U	N	N	U	Y	Y	U	Y	Y	Y	Y
Chen et al. (2025)	Y	U	U	U	U	U	Y	U	U	Y	U	Y	Y

Table 2B. JBI-informed appraisal of intervention and quasi-experimental studies

Table 2B. JBI-informed appraisal of intervention and quasi-experimental studies										
Study	Clear cause/effect	Comparable participants/groups	Similar care apart from intervention	Control/comparator	Multiple measurements	Complete follow-up	Outcomes measured consistently	Reliable outcomes	Appropriate analysis	Principal concern
Laupichler et	Y	NA	NA	N	Y	U	Y	U	Y	No concurrent comparator; self-reported outcome



al. (2022)										
Cheng et al. (2020)	Y	Y	Y	Y	Y	U	Y	Y	Y	Randomization procedure not established from supplied abstract
Wang et al. (2024)	Y	U	U	Y	U	U	Y	Y	Y	Historical cohorts; temporal/cohort confounding
Lyu et al. (2024)	Y	U	Y	Y	Y	U	Y	Y	Y	Very small groups and differing baseline clinical experience
Tabuchi et al. (2024)	Y	NA	NA	N	Y	U	Y	Y	Y	Single-group design; no concurrent control
Lee et al. (2024)	Y	NA	NA	N	Y	Y	Y	Y	Y	Repeated cases and sequential exposure create testing/order effects
Finkelstein et al. (2024)	Y	NA	NA	N	Y	Y	Y	Y	Y	Single-group design; no concurrent comparator
van Kooten et al. (2024)	Y	NA	NA	N	Y	N	Y	U	Y	Different pre/post respondent numbers; predominantly self-report
Kim et al. (2025)	Y	NA	NA	Y	Y	Y	U	Y	Y	Very small sample; order/counterbalancing unclear
Donkor et al. (2025)	Y	NA	NA	N	Y	U	Y	Y	Y	Identical pre/post questions increase testing-effect risk
Stogiannos et al. (2025)	Y	NA	NA	N	Y	N	Y	U	Y	68 pre vs 31 post respondents; self-reported outcomes



Table 2C. Appraisal of implementation, platform, and reader-evaluation studies

Study	Clearly defined population/sample	Intervention/exposure clearly described	Outcome criteria clearly defined	Outcome measurement appropriate	Analysis appropriate	Main methodological limitation
Hernández-Rodríguez et al. (2023)	Y	Y	Y	Y	U	Primarily usability/acceptability; no objective pre/post educational-effectiveness assessment
Bala et al. (2025)	Y	Y	Y	Y	Y	Small resident sample and only 10 confirmed-error test reports
Zhong & Chow (2025)	Y	Y	Y	Y	Y	Primarily perception/acceptability outcomes; limited evidence of objective learning
Verdone et al. (2026)	Y	Y	Y	Y	Y	Small reader sample; evaluates feedback performance/helpfulness rather than longitudinal learning
Chapellier et al. (2026)	Y	Y	Y	Y	Y	Small trainee sample; outcomes predominantly perceived usefulness rather than measured learning gain

Note. Y = Yes; N = No; U = Unclear; NA = Not applicable. “Unclear” indicates that the available study report did not provide sufficient information to support a definitive judgment. Checklist responses were not converted into numerical quality scores. For implementation and reader-evaluation studies that did not map directly onto an intervention checklist, design-relevant methodological domains are presented descriptively.

SUMMARY OF FINDINGS

Across the 17 included studies, AI-assisted educational interventions were generally associated with favourable objective learning, diagnostic-performance, and learner-reported outcomes, although the magnitude and strength of evidence varied according to intervention type, learner population, outcome measure, and study design. Controlled studies generally favoured AI-assisted approaches for image interpretation and diagnostic performance, particularly when AI provided direct image-based assistance, repeated case-based training, or personalised educational feedback. Uncontrolled studies likewise reported improvements in theoretical AI knowledge and image-interpretation performance, although these findings were derived from designs with more limited ability to establish causal effects.

Learner-reported and implementation findings were predominantly favourable, with improvements in confidence, perceived knowledge, usability, educational usefulness, and acceptance of AI-supported instruction. LLM-generated feedback was generally considered useful as a supplement to educator feedback, and some evidence suggested greater perceived benefit among less-experienced trainees. However, concerns were reported regarding AI reliability, hallucinations, clinical fidelity of generated images, data security, accountability, and overreliance on technology. Overall, the findings indicate potential educational benefits across several forms of AI-assisted medical-imaging education, while substantial methodological and clinical heterogeneity limits direct comparison across studies.

DISCUSSION

This systematic review synthesized evidence from 17 studies evaluating AI-assisted educational interventions across medical-imaging education. Overall, the findings suggest that AI-assisted approaches have potential to improve objective learning and diagnostic performance while also supporting learner confidence, engagement, and perceived educational usefulness. The strongest evidence of educational effectiveness came from controlled studies evaluating image interpretation, diagnostic performance, and personalized feedback,



whereas uncontrolled studies generally reported improvements in AI knowledge and diagnostic competency. Studies focused on implementation and learner experience also showed predominantly favorable perceptions, although concerns regarding reliability, hallucinations, clinical fidelity, accountability, data security, and overreliance on AI remained. Importantly, the considerable heterogeneity in learner populations, intervention types, assessment methods, and study designs limits direct comparison and indicates that the educational value of AI is likely to depend on how, where, and for whom it is implemented.

Effects on Objective Learning and Diagnostic Competency

A central finding of this review was the generally favorable effect of AI-assisted interventions on objective learning and diagnostic outcomes. Controlled studies provided particularly encouraging evidence. Cheng et al. (2020), for example, demonstrated that medical students receiving AI-assisted hip-fracture interpretation training achieved a significantly greater improvement in diagnostic accuracy than students receiving conventional learning. The persistence of improved performance during subsequent unaided testing is noteworthy because it suggests that the intervention may have supported learning rather than merely improving performance while AI assistance was available. The greater benefit observed among initially lower-performing students also raises the possibility that appropriately designed AI tools may help address differences in baseline competency by providing targeted visual support during learning.

Similar findings were observed across other imaging applications. Gokkurt Yilmaz et al. (2025) found that AI-generated personalized feedback produced greater improvement in dental-radiograph diagnostic performance than conventional feedback, while Wang et al. (2024) reported higher practical, CT, and MRI diagnostic scores among medical students receiving AI-assisted instruction. Lyu et al. (2024) further demonstrated improvements in pulmonary-nodule interpretation during repeated AI-assisted training. Collectively, these findings suggest that AI may be particularly useful when incorporated directly into active diagnostic learning rather than presented only as theoretical content.

The findings of Chen et al. (2025) provide an additional perspective on the role of AI relative to conventional educator-led instruction. AI-assisted thyroid-ultrasound training produced greater diagnostic accuracy than conventional training alone but did not significantly outperform additional specialty training. This finding is important because the educational objective of AI may not necessarily be to surpass expert instruction. Instead, AI could provide an additional mechanism for repeated practice, feedback, and structured learning when continuous specialist supervision is difficult to provide. In this context, comparable performance between AI-assisted and specialty training may itself represent an educationally meaningful outcome.

Nevertheless, improvement was not restricted to controlled studies. Donkor et al. (2025) and Finkelstein et al. (2024) reported substantial increases in objective AI knowledge following structured educational programmes, while Kim et al. (2025), Lee et al. (2024), and Tabuchi et al. (2024) demonstrated improvements in image interpretation or diagnostic performance following LLM-supported learning, digital self-learning with AI assistance, and synthetic-image training, respectively. Particularly notable was the improvement reported by Tabuchi et al. (2024) on previously unseen standard-field retinal images after learners had trained using AI-generated ultra-widefield images. This suggests that appropriately designed synthetic-image education may facilitate transfer of acquired recognition skills beyond the exact image format encountered during training.

These uncontrolled findings should, however, be interpreted more cautiously. Without parallel comparator groups, observed improvements cannot be attributed exclusively to AI-assisted education. Repeated testing, exposure to the same or similar cases, conventional learning occurring concurrently with the intervention, and familiarity with assessment procedures may contribute to observed gains. The objective improvements across these studies are therefore encouraging but require confirmation through larger controlled trials.

Educational Functions of Different AI Approaches

The interventions included in this review demonstrate that AI-assisted medical-imaging education is not a single educational strategy. Rather, AI served several distinct functions, including diagnostic decision support, automated feedback, personalized learning, generation of educational images, simulation, information



retrieval, and structured instruction in AI concepts. This distinction is important because the educational effectiveness of an intervention may depend more on its pedagogical function than on the specific AI technology employed.

AI-assisted image interpretation was among the approaches most consistently associated with objective performance improvements. Tools providing visual cues, lesion annotations, diagnostic suggestions, or computer-assisted diagnosis allowed learners to compare their reasoning with algorithm-supported information during image interpretation. Such approaches were represented in hip-fracture detection, pulmonary-nodule assessment, thyroid ultrasound, CT and MRI interpretation, and retinal disease recognition. The consistency of favorable findings across different imaging applications suggests that AI may be particularly suited to educational tasks requiring repeated pattern recognition and interpretation of large numbers of images.

Personalization and feedback represented another potentially important educational function. The findings of Gokkurt Yilmaz et al. (2025) suggest that AI-generated learning materials tailored to individual knowledge gaps may be more effective than generic correct/incorrect feedback. Similarly, Bala et al. (2025) and Verdone et al. (2026) demonstrated that LLMs could identify discrepancies or errors in radiology reports and generate feedback that learners generally considered useful. These applications may be particularly relevant in competency-based education, where learners require timely and specific feedback on individual weaknesses.

However, the evidence for automated feedback remains more convincing in terms of feasibility and acceptability than demonstrated learning effectiveness. Neither Bala et al. (2025) nor Verdone et al. (2026) established that receiving LLM-generated feedback subsequently improved residents' independent reporting or diagnostic performance. The ability of an AI system to identify errors accurately should therefore not be considered equivalent to evidence that the system improves learning. Future evaluations should extend beyond technical accuracy and learner satisfaction to determine whether repeated exposure to AI-generated feedback produces measurable and retained improvements in competency.

Generative AI also introduced educational possibilities beyond conventional decision-support systems. Tabuchi et al. (2024) demonstrated the potential of generative AI to create large numbers of synthetic images for diagnostic training, whereas Zhong and Chow (2025) explored generative-AI simulation for radiography education. Synthetic data could potentially increase learner exposure to diverse cases, including uncommon findings that may be difficult to encounter consistently during clinical placement. Nevertheless, the concerns regarding image realism and clinical fidelity reported by Zhong and Chow (2025) emphasize that generated educational material requires careful validation before being used for higher-level diagnostic training.

Large Language Models in Medical-Imaging Education

The emergence of LLM-based interventions was particularly evident among the more recent studies. LLMs were used for personalized feedback, diagnostic information retrieval, automated reporting feedback, and generation of differential diagnoses. Kim et al. (2025) showed that a brief structured tutorial on using an LLM-based search tool improved residents' diagnostic performance compared with both untrained LLM use and conventional internet searching. This finding suggests that educational outcomes may depend not only on access to AI but also on whether learners are taught how to use it effectively.

This has important implications for curriculum design. Simply providing learners with generative-AI tools may not be sufficient to achieve educational benefits. Learners may require instruction in prompt formulation, verification of generated information, recognition of uncertainty, and critical appraisal of AI outputs. The hallucination rate reported by Kim et al. (2025), although relatively low, demonstrates that plausible but incorrect information remains a relevant educational risk. If learners lack sufficient foundational knowledge to recognize such errors, AI assistance could potentially reinforce misconceptions rather than correct them.

The findings of Chapellier et al. (2026) further suggest that learner experience may influence the perceived value of LLM-supported education. Junior radiology trainees were substantially more likely than advanced trainees to rank LLM-generated teaching as their preferred educational source. This may indicate that less-experienced learners benefit more from structured differential diagnoses and readily accessible explanations,



whereas advanced trainees rely more heavily on established clinical knowledge and experience. AI-assisted education may therefore need to be adapted to learner level rather than applied uniformly throughout training.

Learner Confidence, Acceptability, and Implementation

The implementation-focused studies generally demonstrated favorable learner attitudes toward AI-assisted education. Laupichler et al. (2022), Stogiannos et al. (2025), and Van Kooten et al. (2024) reported improvements in perceived AI readiness, knowledge, practical skills, or confidence following relatively short educational interventions. These findings suggest that structured exposure to AI may help learners become more comfortable with technologies that are increasingly encountered in imaging practice.

Nevertheless, improvements in self-perceived knowledge or confidence should not be interpreted as equivalent to improvements in objective competency. Several implementation studies lacked objective knowledge assessments, and some had substantial differences between pre-intervention and post-intervention response numbers. Confidence may be educationally valuable, particularly when introducing unfamiliar technology, but excessive confidence without corresponding competency could create additional risks. Future educational evaluations should therefore measure subjective and objective outcomes concurrently to determine whether increased confidence is supported by measurable improvements in knowledge or performance.

Learners also expressed several recurring concerns. These included AI reliability, hallucinations, accountability, data security, patient confidentiality, technological overreliance, and the clinical realism of AI-generated images. Such concerns demonstrate that successful implementation requires more than demonstrating educational effectiveness. Medical-imaging programmes need to consider technological infrastructure, governance, privacy, educator oversight, and mechanisms for validating AI-generated educational content.

The preference reported by Bala et al. (2025) for combining LLM-generated feedback with traditional educator feedback is particularly relevant. Rather than supporting complete replacement of educators, the evidence suggests that learners may value AI most as an additional source of timely feedback that complements human expertise. Educators remain important for contextualizing AI outputs, addressing complex reasoning, identifying inappropriate recommendations, and helping learners understand the clinical significance of errors.

Methodological Considerations and Strength of the Evidence

Although the overall direction of findings was favorable, the methodological characteristics of the included literature require cautious interpretation. Sample sizes varied considerably, and several studies included relatively few learners. Kim et al. (2025), for example, evaluated only nine radiology residents, while Lyu et al. (2024) included six participants in each study group. Small samples reduce statistical precision and may limit generalizability to broader learner populations.

Several studies also lacked parallel control groups, while others used historical controls or within-participant comparisons. Repeated assessment of identical or similar cases may introduce recall, learning, or ordering effects. In pre-test/post-test designs, improvements may partly reflect familiarity with the assessment rather than the intervention itself. Self-selection and differential response rates may further influence studies relying on voluntary participation or pre- and post-intervention surveys.

Outcome heterogeneity represents another major limitation of the evidence base. Studies assessed knowledge scores, diagnostic accuracy, AUC, kappa statistics, confidence, readiness, satisfaction, usability, and perceived usefulness using different instruments and follow-up periods. Intervention duration ranged from brief tutorials or seminars to programmes extending several months. This heterogeneity makes direct comparison difficult and limits the appropriateness of quantitative pooling.

Long-term educational outcomes were also insufficiently examined. Most studies evaluated performance immediately or relatively soon after the intervention. Gokkurt Yilmaz et al. (2025) included a one-month reassessment, but evidence regarding longer-term retention, transfer into clinical practice, and effects on



independent professional performance remained limited. Short-term improvement in test performance does not necessarily indicate durable competency or improved clinical decision-making.

Implications for Medical-Imaging Education

The findings have several implications for educators and curriculum developers. First, AI-assisted education appears most promising when it is integrated into an identifiable educational activity with clear learning objectives rather than introduced simply as access to a technological tool. Interventions involving repeated image interpretation, targeted feedback, personalized learning, and structured AI-use training produced some of the most favorable findings in this review.

Second, AI education should include both learning with AI and learning about AI. As AI becomes increasingly integrated into clinical imaging, learners need practical experience interpreting AI-supported information while also understanding its limitations. Curricula should therefore address algorithmic errors, hallucinations, bias, data governance, privacy, accountability, and the need to verify AI-generated recommendations.

Third, implementation should preserve meaningful educator involvement. The evidence does not establish that AI should replace conventional instruction, specialist teaching, or educator feedback. Instead, AI may expand opportunities for practice and provide more immediate or individualized support between educator interactions. This may be particularly valuable in settings where faculty time, specialist supervision, or access to diverse clinical cases is limited.

The findings may also have relevance for institutions with varying technological resources. Some interventions required sophisticated AI-CAD systems or large clinical datasets, whereas others used commercially available LLMs, online modules, or relatively brief structured curricula. Consequently, AI-assisted education does not necessarily require a single high-resource implementation model. Institutions may need to select approaches according to curricular objectives, infrastructure, faculty preparedness, learner access, governance requirements, and the clinical relevance of available technologies.

FUTURE RESEARCH

Future research should prioritize adequately powered multicentre randomized or well-controlled studies with clearly defined educational interventions and validated objective outcome measures. Studies should distinguish between improvement occurring while AI assistance is actively available and independent performance after the tool has been removed. This distinction is essential for determining whether AI produces genuine learning rather than temporary performance augmentation.

Longer follow-up periods are also needed to determine whether improvements in knowledge and diagnostic performance are retained and transferred to new cases, different imaging modalities, and clinical practice. Assessments using previously unseen cases should be encouraged to reduce recall effects and provide stronger evidence of generalizable learning. Where possible, studies should compare AI-assisted instruction not only with no intervention but also with established educator-led approaches.

Future studies of LLM-generated feedback should evaluate whether repeated use improves independent diagnostic reasoning and reporting performance rather than focusing primarily on the technical quality or perceived usefulness of generated feedback. Research should also examine hallucinations, automation bias, overreliance, and whether learners at different stages of training respond differently to AI assistance.

Finally, implementation research should examine feasibility across different educational and geographic settings, particularly institutions with limited technological infrastructure or faculty expertise. Cost, accessibility, educator workload, data privacy, governance, learner equity, and sustainability should be considered alongside educational effectiveness. Establishing standardized outcome measures for AI-assisted medical-imaging education would further improve comparability across studies and strengthen future evidence synthesis.



STRENGTHS AND LIMITATIONS OF THE REVIEW

A strength of this systematic review is its broad consideration of AI-assisted education across multiple medical-imaging disciplines, learner populations, and AI technologies. The review distinguished objective learning and diagnostic outcomes from subjective and implementation outcomes and separated controlled from uncontrolled evidence, reducing the risk of interpreting learner satisfaction or confidence as equivalent to demonstrated educational effectiveness. It also considered emerging applications including LLM-generated feedback, personalized learning, AI-assisted image interpretation, and generative-AI synthetic imaging.

Several limitations should nevertheless be acknowledged. The evidence base was highly heterogeneous in study design, participant characteristics, intervention type, duration, comparator, and outcome measurement, limiting direct comparison and quantitative synthesis. Several included studies had small samples, lacked control groups, relied on self-reported outcomes, or assessed only short-term effects. Restriction to English-language publications may also have excluded relevant evidence published in other languages. In addition, the rapid evolution of AI technologies means that newer educational applications may emerge faster than they can be captured through conventional systematic-review processes. These limitations should be considered when interpreting the overall findings and their generalizability.

CONCLUSION

AI-assisted educational interventions show promising potential to enhance medical-imaging education, particularly in image interpretation, diagnostic performance, personalized learning, and provision of timely educational feedback. Controlled evidence generally favored AI-assisted approaches, while uncontrolled studies also demonstrated improvements in knowledge and diagnostic competency. Learners generally perceive AI-supported education positively, although concerns regarding reliability, hallucinations, clinical fidelity, privacy, accountability, and technological overreliance remain important. Current evidence supports AI primarily as an adjunct to structured educator-led learning rather than a replacement for educators or established teaching methods. Its educational value appears to depend on appropriate instructional design, learner preparation, validation of AI-generated content, and continued human oversight. Larger controlled studies with standardized objective outcomes, longer follow-up, and assessment of independent clinical performance are needed to establish the durability, generalizability, and implementation readiness of AI-assisted education across medical-imaging disciplines.

Generative AI Declaration

During the preparation of this manuscript, the authors used ChatGPT for proofreading assistance. All content was subsequently reviewed and edited by the authors, who assume full responsibility for the accuracy and integrity of the published work.

REFERENCES

1. Arif, W. M. (2024). Radiologic technology students' perceptions on adoption of artificial intelligence technology in radiology. *International Journal of General Medicine*, 17, 3129–3136. <https://doi.org/10.2147/IJGM.S465944>
2. Bala, W., Li, H., Moon, J., Trivedi, H., Gichoya, J., & Balthazar, P. (2025). Enhancing radiology training with GPT-4: Pilot analysis of automated feedback in trainee preliminary reports. *Current Problems in Diagnostic Radiology*, 54(2), 151–158. <https://doi.org/10.1067/j.cpradiol.2024.08.003>
3. Biswas, S. (2023). Artificial intelligence in radiology education. *International Journal of Radiology & Radiation Therapy*, 10(1), 1–3. <https://doi.org/10.15406/ijrrt.2023.10.00344>
4. Cacho, R. (2024). Integrating generative AI in university teaching and learning: A model for balanced guidelines. *Online Learning*, 28(3), 55–81. <https://doi.org/10.24059/olj.v28i3.4508>
5. Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368, l6890.



<https://doi.org/10.1136/bmj.l6890>

6. Cayabyab, M. A. S., Rejollo, M. R. A., Rayco, R. C. M., Laping, I. P., Tan, J. A., Castañares, L. D. W., Castro, A. M. L., Melford, C. M. (2026). Impact of AI-Assisted Chest Radiography as a Triage Tool for Molecular Lab Testing in Rural Philippines: A Systematic Review of Diagnostic Yield and Health Equity. *International Journal of Research and Innovation in Social Science*, 10(6), <https://doi.org/10.47772/IJRISS.2026.100600430>
7. Chapellier, P., Ferrari, J., Saliba, T., Jeltsch, P., Mohamed, M., Jankovski, S., Ilanjan, G., Epis, M., Pansini, V., Bragaglia, F., Agostinelli, A., Caringal, K., Lalov, L., Rotzinger, D. C., & Fahrni, G. (2026). Large language model-generated differential diagnoses in radiology education: Comparison with a standard casebook. *Diagnostics*, 16(13), 2009. <https://doi.org/10.3390/diagnostics16132009>
8. Chau, M., Arruzza, E., & Johnson, N. (2022). Simulation-based education for medical radiation students: A scoping review. *Journal of Medical Radiation Sciences*, 69(3), 367–381. <https://doi.org/10.1002/jmrs.572>
9. Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
10. Chen, Y., Sun, Z., Lin, W., Xv, Z., & Su, Q. (2025). Artificial intelligence in the training of radiology residents: A multicenter randomized controlled trial. *Journal of Cancer Education*, 40(2), 234–240. <https://doi.org/10.1007/s13187-024-02502-0>
11. Cheng, C.-T., Chen, C.-C., Fu, C.-Y., Chaou, C.-H., Wu, Y.-T., Hsu, C.-P., Chang, C.-C., Chung, I.-F., Hsieh, C.-H., Hsieh, M.-J., & Liao, C.-H. (2020). Artificial intelligence-based education assists medical students' interpretation of hip fracture. *Insights into Imaging*, 11, 119. <https://doi.org/10.1186/s13244-020-00932-0>
12. Co, S.J. (2025). Artificial intelligence in Philippine education: A narrative review of applications, perceptions, and challenges. *International Research Journal of Science, Technology, Education, and Management*, 5(2), 25-38. <https://doi.org/10.5281/zenodo.15902950>
13. Crotty, E., Singh, A., Neligan, N., Chamunyonga, C., & Edwards, C. (2024). Artificial intelligence in medical imaging education: Recommendations for undergraduate curriculum development. *Radiography*, 30, 67–73. <https://doi.org/10.1016/j.radi.2024.10.012>
14. Doherty, G., McLaughlin, L., Hughes, C., McConnell, J., Bond, R., & McFadden, S. (2024). A scoping review of educational programmes on artificial intelligence (AI) available to medical imaging staff. *Radiography*, 30(2), 474–482. <https://doi.org/10.1016/j.radi.2023.12.019>
15. Donkor, A., Boakye, E., Atuanor, P., & Wiafe, Y. A. (2025). Evaluation of a classroom-based medical imaging artificial intelligence educational intervention in Ghana: A pre-test/post-test study design. *Radiography*, 31(4), 102987. <https://doi.org/10.1016/j.radi.2025.102987>
16. du Plooy, E., Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), e39630. <https://doi.org/10.1016/j.heliyon.2024.e39630>
17. Duong, M. T., Rauschecker, A. M., Rudie, J. D., Chen, P.-H., Cook, T. S., Bryan, R. N., & Mohan, S. (2019). Artificial intelligence for precision education in radiology. *British Journal of Radiology*, 92(1103), 20190389. <https://doi.org/10.1259/bjr.20190389>
18. Finkelstein, M., Ludwig, K., Kamath, A., Halton, K. P., & Mendelson, D. S. (2024). The impact of an artificial intelligence certificate program on radiology resident education. *Academic Radiology*, 31(11), 4709–4714. <https://doi.org/10.1016/j.acra.2024.05.041>
19. Garin, S. P., Zhang, V., Jeudy, J., Parekh, V. S., & Yi, P. H. (2023). Systematic review of radiology residency artificial intelligence curricula: Preparing future radiologists for the artificial intelligence era. *Journal of the American College of Radiology*, 20(6), 561–569. <https://doi.org/10.1016/j.jacr.2023.02.031>
20. Gokkurt Yilmaz, B. N., Ozbey, F., & Yilmaz, B. E. (2025). Effect of artificial intelligence-assisted personalized feedback on radiographic diagnostic performance of dental students: A controlled study. *BMC Medical Education*, 25, 1403. <https://doi.org/10.1186/s12909-025-07875-4>
21. Hernández-Rodríguez, J., Rodríguez-Conde, M.-J., Santos-Sánchez, J.-Á., & Cabrero-Fraile, F.-J. (2023). Development and validation of an educational software based in artificial neural networks for training in radiology (JORCAD) through an interactive learning activity. *Heliyon*, 9(4), e14780. <https://doi.org/10.1016/j.heliyon.2023.e14780>



22. International Society of Radiographers and Radiological Technologists. (ISRRT, 2004, November). Guidelines for the education of entry-level professional practice in medical radiation sciences. https://www.isrtr.org/wp-content/uploads/2023/07/Document_6_Standards_of_Education.pdf
23. Keshavarz, P., Mohammadigoldar, Z., Bedayat, A., Raman, S. S., & Tai, R. (2026). Artificial intelligence education in radiology training: A systematic review of effectiveness, barriers, and future directions. *Academic Radiology*, 33(3), 695–706. <https://doi.org/10.1016/j.acra.2025.10.049>
24. Kim, S. H., Schramm, S., Wihl, J., Raffler, P., Tahedl, M., Canisius, J., Luiken, I., Endrös, L., Reischl, S., Marka, A., Walter, R., Schillmaier, M., Zimmer, C., Wiestler, B., & Hedderich, D. M. (2025). Boosting LLM-assisted diagnosis: 10-minute LLM tutorial elevates radiology residents' performance in brain MRI interpretation. *Neuroradiology*, 67(8), 2069–2081. <https://doi.org/10.1007/s00234-025-03664-4>
25. Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
26. Lai, N. M., Lim, Y. S., Win, M. T., Bhargava, P., Thomas, P., & Ong, Q. C. (2026). The effectiveness of artificial intelligence in undergraduate health professions education: Systematic review and meta-analysis of randomized controlled trials. *JMIR Medical Education*, 12, e88933. <https://doi.org/10.2196/88933>
27. Laupichler, M. C., Hadizadeh, D. R., Wintergerst, M. W. M., von der Emde, L., Paech, D., Dick, E. A., & Raupach, T. (2022). Effect of a flipped classroom course to foster medical students' AI literacy with a focus on medical imaging: A single group pre-and post-test study. *BMC Medical Education*, 22, Article 803. <https://doi.org/10.1186/s12909-022-03866-x>
28. Lee, S. E., Kim, H. J., Jung, H. K., Jung, J. H., Jeon, J.-H., Lee, J. H., Hong, H., Lee, E. J., Kim, D., & Kwak, J. Y. (2024). Improving the diagnostic performance of inexperienced readers for thyroid nodules through digital self-learning and artificial intelligence assistance. *Frontiers in Endocrinology*, 15, 1372397. <https://doi.org/10.3389/fendo.2024.1372397>
29. Létourneau, A., Deslandes Martineau, M., Charland, P., Karan, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *npj Science of Learning*, 10, 29. <https://doi.org/10.1038/s41539-025-00320-7>
30. Lewis, S., Bhyat, F., Casmod, Y., Gani, A., Gumede, L., Hajat, A., Hazell, L., Kammies, C., Mahlaola, T. B., Mokoena, L., & Vermeulen, L. (2024). Medical imaging and radiation science students' use of artificial intelligence for learning and assessment. *Radiography*, 30(Suppl. 2), 60–66. <https://doi.org/10.1016/j.radi.2024.10.006>
31. Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 22, 42. <https://doi.org/10.1186/s41239-025-00540-2>
32. Lyu, X., Dong, L., Fan, Z., Sun, Y., Zhang, X., Liu, N., & Wang, D. (2024). Artificial intelligence-based graded training of pulmonary nodules for junior radiology residents and medical imaging students. *BMC Medical Education*, 24, 740. <https://doi.org/10.1186/s12909-024-05723-5>
33. Masters, K., MacNeil, H., Benjamin, J., Carver, T., Nemethy, K., Valanci-Aroesty, S., Taylor, D. C. M., Thoma, B., & Thesen, T. (2025). Artificial intelligence in health professions education assessment: AMEE Guide No. 178. *Medical Teacher*, 47(9), 1410–1424. <https://doi.org/10.1080/0142159X.2024.2445037>
34. Melford, C. M., O'Brien-Melford, M. A. P. (2026). Artificial Intelligence Applications in Clinical Laboratory Diagnostics: a Systematic Review of Diagnostic Accuracy, Workflow Efficiency, and Clinical Utility. *International Journal of Research and Scientific Innovation (IJRSI)*, 13(4), <https://doi.org/10.51244/IJRSI.2026.1304000242>
35. 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., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>



36. Salastekar, N. V., Maxfield, C., Hanna, T. N., Krupinski, E. A., Heitkamp, D. E., & Grimm, L. J. (2023). Artificial intelligence/machine learning education in radiology: Multi-institutional survey of radiology residents in the United States. *Academic Radiology*, 30(7), 1481–1487. <https://doi.org/10.1016/j.acra.2023.01.005>
37. Suazo-Galdamés, I. C., & Chaple-Gil, A. M. (2025). AI-powered adaptive learning systems in higher education: A scoping review of implementation and impact on academic performance. *Data and Metadata*, 4, 981. <https://doi.org/10.56294/dm2025981>
38. Stogiannos, N., Skelton, E., Kumar, S., Ahmed, S., Amedu, C., Vince, C., Schiavottiello, M., O'Sullivan, C., & Malamateniou, C. (2025). Evaluation of a customised, AI-focused educational seminar delivered to final year undergraduate radiography students in the UK: A cross-sectional study. *Radiography*, 31(3), 102926. <https://doi.org/10.1016/j.radi.2025.102926>
39. Tabuchi, H., Engelmann, J., Maeda, F., Nishikawa, R., Nagasawa, T., Yamauchi, T., Tanabe, M., Akada, M., Kihara, K., Nakae, Y., Kiuchi, Y., & Bernabeu, M. O. (2024). Using artificial intelligence to improve human performance: Efficient retinal disease detection training with synthetic images. *British Journal of Ophthalmology*, 108(10), 1430–1435. <https://doi.org/10.1136/bjo-2023-324923>
40. UNESCO. (2023). Global education monitoring report 2023: Technology in education—A tool on whose terms? United Nations Educational, Scientific and Cultural Organization. <https://doi.org/10.54676/UZQV8501>
41. United Nations Educational, Scientific and Cultural Organization. (UNESCO, 2025). Philippines: Artificial intelligence readiness assessment report. <https://unesdoc.unesco.org/ark:/48223/pf0000393860>
42. 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, 15. <https://doi.org/10.1186/s13244-023-01595-3>
43. Verdone, A., Cardall, A., Siddiqui, F., Nashawaty, M., Rigau, D., Kwon, Y., Yousef, M., Patel, S., Kieturakis, A., Kim, E., Heacock, L., Reig, B., & Shen, Y. (2026). Evaluating generative artificial intelligence as an educational tool for radiology resident report drafting. *Journal of the American College of Radiology*, 23(5), 818–826. <https://doi.org/10.1016/j.jacr.2025.12.024>
44. Vergara, J. P. C. (2025). Navigating the generative artificial intelligence era: Charting the course for curricular reform in higher education in the Philippines. Second Congressional Commission on Education. <https://edcom2.gov.ph/publications/navigating-the-generative-artificial-intelligence-era-charting-the-course-for-curricular-reform-in-higher-education-in-the-philippines/>
45. Villarino, R. T. (2025). Artificial Intelligence (AI) integration in Rural Philippine Higher Education: Perspectives, challenges, and ethical considerations. *IJERI: International Journal of Educational Research and Innovation*, (23). <https://doi.org/10.46661/ijeri.10909>
46. Wang, D., Huai, B., Ma, X., Jin, B., Wang, Y., Chen, M., Sang, J., & Liu, R. (2024). Application of artificial intelligence-assisted image diagnosis software based on volume data reconstruction technique in medical imaging practice teaching. *BMC Medical Education*, 24, Article 405. <https://doi.org/10.1186/s12909-024-05382-6>
47. Wang, X., Huang, R., Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024). The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis. *Journal of Educational Computing Research*, 62(6), 1568–1603. <https://doi.org/10.1177/07356331241240459>
48. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
49. Zhong, D., & Chow, S. K. K. (2025). Investigating the potential of generative AI clinical case-based simulations on radiography education: A pilot study. *Journal of Imaging Informatics in Medicine*, 39(2), 1848–1860. <https://doi.org/10.1007/s10278-025-01601-8>