

A Critical Narrative Synthesis Toward Precision Medical Training in Radiology Education

Kenan Alkhalili, M.D.^{1a}, Sarah Ali, M.D.²

¹ Department of Radiology, University of California, San Francisco, ² Department of Radiology, Brigham and Women's Hospital

<https://doi.org/10.64551/001c.150391>



Inquisiva Open

Vol. 1, Issue 1, 2025

Abstract

Objectives

The research examines the intersection of artificial intelligence (AI), including large language models (LLMs), generative platforms, and adaptive simulations, with radiology education as a testing ground for precision learning, curricular creativity, faculty preparation, and governance. The aims here are not to catalog all of the uses, but to critically examine convergent findings and to comment on the mechanisms by which AI enhances, or subtracts from, diagnostic training.

Methods

Critical narrative synthesis guided by design-oriented perspective with abductive coding that integrated theory-driven and inductive approaches. Evidence was extracted using iterative peer-reviewed database searching and tracing of review-worthy reviews, studies, and society guidelines. Sources were kept that discussed AI application in radiology teaching and presented mechanisms of effect, recorded results, or protection measures against irresponsible adoption.

Findings

Three common mechanisms across various studies emerge: adaptive case curation that broadens exposure to underrepresented pathology; iterative feedback loops that move assessment from episodic to longitudinal; and language scaffolding that improves report clarity and professional communication. Large language models apply these mechanisms further to curriculum design, interactive teaching, and assessment design, while gamification and simulation provide depth of experience and fun. Those advantages come with risks of overdependence, diagnostic deskilling, bias, and misinformation that are significant with low faculty readiness or lack of structures of oversight. Protections of human-in-the-loop frameworks, alternating between AI-assistance and AI-restricted interpretation, and redesign of assessments to reward independent reasoning are always urged. Curricular adoption worldwide is uneven, with interest among trainees frequently exceeding that of faculty, and with an associated generational divide. Professional organizations released initial recommendations, but agreement on standards of readiness and mechanisms of governance is partial.

Conclusions

Radiology exemplifies that with AI, precision medical education can be taken to the next level, but the gains depend on design. The AI should be integrated not to replace expertise but to facilitate intentional practice, critical thought, and access equity. The integration

^a Corresponding Author:

Kenan Alkhalili M.D., kenan_alkhalili@hotmail.com

Department of Radiology, University of California, San Francisco

provides design postulations, research agendas, and practice-focused implications to inform responsible adoption, and radiology can both act as a template and an exemplar case to apply to wider medical education.

Keywords: Radiology education, Generative AI, Medical education, Curriculum design, Programmatic assessment, Radiology

1. Introduction

Artificial intelligence shifted promptly from speculation to practice in health care, and with that, opportunities and fears were raised about medical education. Nowhere is this clearer than with radiology, a discipline based on pipelined data structures, digital image repositories, and standardized reporting. Unlike medical specialties whose learning experiences rely primarily on bedside or procedure-based interactivity, radiology is particularly well suited to computational enhancement. Through its rich informatics tradition and heavy dependence on visual pattern recognition, radiology is an ideal proving ground for pedagogy with AI.

In this dynamic context, large language models (LLMs), generative artificial intelligence (ai) systems, and adaptive simulations emerge as particularly promising to revolutionize the way knowledge is learned, cases are assembled, and feedback is provided. The tools can now not only extract information retrieved but can author reports, mimic diagnostic reasoning, construct custom study cases, and even provide criticism of outputs by trainees. Their ability to scale feedback and diversify case exposure has made them potential augmentative counterparts to conventional education. However, their stochastic character, vulnerability to bias, and relative lack of transparency all concern reliability, ethics, and equity. The need for critical synthesis arises out of a basic dilemma: while AI offers remedies to perennial problems of case exposure variability and scarce opportunities for faculty time, it carries with it associated risks that can harm independent diagnostic reasoning. The radiology discipline, by virtue of its maturity of data infrastructure, falls prey to this dualism more dramatically than many others, and initial accounts have emphasized the need for formalized curricula of AI, foundational training material, and tracks of precision education (Tejani et al., 2022). Understanding how AI redesigns learning mechanisms, and what safeguarding is necessary to forestall its misapplication, necessitates not cataloging of uses; rather, theory-influenced analyses of mechanisms, of risks, and of conditions of proper deployment.

This paper therefore pursues four guiding research questions:

1. RQ1: Which persistent learning problems in radiology education are most plausibly addressable by AI-enabled tools?
2. RQ2: Through what mechanisms do AI tools, including LLMs, generative platforms, and adaptive simulations, alter diagnostic learning processes such as case exposure, feedback, and reasoning?
3. RQ3: What safeguards are required to ensure that AI functions as a scaffold for expertise rather than a substitute that risks deskilling, bias, or invalid assessment?

4. RQ4: Under what conditions (faculty readiness, professional perceptions, governance structures, and equity considerations) does integration of AI in radiology education become educationally defensible?

Hence, this paper aims to clarify radiology's position not only as an early adopter of AI in education but also as a model for how precision medical education might evolve. The sections that follow outline the synthesis method, analyze learning mechanisms and engagement strategies, examine risks and safeguards, and consider implications for curriculum design, faculty readiness, professional culture, and future research priorities.

2. Methods

The report takes on a critical narrative synthesis with a design-oriented perspective. The goal here is not to produce a systematic or scoping review that is comprehensive but to combine converging theory and evidence to inform us about how radiology education is being transformed by artificial intelligence (AI) technologies, how they pose a threat, and in which instances their integration is educationally justified.

Relevant literature was determined by iterative searching of Scopus, PubMed, and Google Scholar between 2018 and 2025 with terms that combined radiology education, artificial intelligence, large language models, generative AI, curriculum, training, and assessment. Tracing of bibliographical references of high-impact reviews and consensus statements was undertaken to determine other studies. Highest priority was given to (a) empirical tests of AI tools within radiology education; (b) umbrella and systematic reviews; (c) program reports and consensus guidelines; and (d) perspective articles that made testable statements about mechanisms, safeguards, or implementation. Opinion pieces without educational implications were excluded unless they made direct comment about issues of governance, ethics, or readiness. Studies were retained if they addressed at least one of the following:

1. AI applications in radiology training or assessment.
2. Documented outcomes related to knowledge acquisition, diagnostic reasoning, or professional readiness.
3. Theoretical or empirical analysis of risks such as overreliance, deskilling, misinformation, or bias.
4. Safeguards and governance strategies relevant to educational integration.

The evidence was abductively coded using theory-driven and inductive strategies. The synthesis was guided by four lenses of analysis: (1) radiology training learning problems (e.g., lack of case diversity, scarcity of feedback, unequal exposure); (2) AI mechanisms of effect (e.g., adaptive cu-

rating of cases, language scaffolds, gamified learning, curriculum augmentation); (3) protection and risk avoidance (e.g., human-in-the-loop approaches, over-reliance avoidance, validity of design of assessments); and (4) conditions of implementation (faculty preparation, integration with curricula, professional attitudes, governance, and equity). The findings were then synthesized into a conceptual model that maps mechanisms to needs, points of risk, and conditions of responsible adoption.

3. Synthesis of Evidence

3.1. Radiology as a Pioneer Discipline

Radiology has long been recognized as one of the medical specialties most receptive to technological innovation, and its position at the vanguard of artificial intelligence (AI) integration into education is not accidental. Various structural attributes render radiology particularly well suited to experimentation with pedagogies powered by artificial intelligence. They comprise a matured informatics culture, an ingrained digital workflow dependence, and decades of familiarity with structured reporting systems that stress data integrity and reproducibility. All of these together provide an environment that can pilot AI tools without putting aside the very substance of clinical practice, and with opportunities and risks coming into sharp and uncommon focus.

Where procedural specialties depend on the tactile and experiential experience of taking care of patients during training, radiology's educational architecture always relied crucially on case collections, digital image libraries, and standardized schemas of interpretation. These resources align naturally with the affordances of AI and specifically with large language models (LLMs) and generative models. Data pipelines that are properly structured allow model training and testing while digital case collections map onto adaptive case curation. The very practice of radiology education (pattern identification, formatted report writing, and comparison of features across images) is cognitively congruent with the abilities of AI to extract patterns and to generate text. That congruence accounts for why radiology came to represent a "proving ground" for educational AI while other specialties continue to ponder questions of viability. Foundational viewpoints have correspondingly highlighted how radiology's infrastructure makes the specialty rich ground on which to launch introductions to AI and initiatives to enhance precision education with AI (Tejani et al., 2022).

The empirical literature mirrors this positioning. Sacransky et al. (2024) and Temperley et al. (2024) report extensive uses of LLMs in purposes ranging from report writing to exam preparation and show how already, AI tools are integral to everyday aspects of the learning experience. Lastrucci et al. (2024) and Keshavarz et al. (2024) carry these observations further by illustrating how generative models are being integrated into virtual reading rooms, organized teaching files, and interactive sites offering iterative practice opportunities to trainees. These reviews converge on the note that the pedagogical value of AI is not

primarily in automating outcomes, but rather in supporting learning processes: report drafts become targets of review, generated cases bridge sparse clinical exposure, and adaptive feedback loops eliminate lag between learner error and corrective guidance.

Meanwhile, this earlier experimentation has revealed tensions that run to the heart of the wider question of AI and medical education. The stochastic quality of LLMs, their unpredictably reliable output, and their capacity to mislead novices are particularly evident within radiology, where errors in interpretation can have immediate implications both for education and care. Radiology therefore exemplifies the dilemma of early adoption: the very qualities that make the specialty an exemplary laboratory of innovation enhance correspondingly the danger of too early or blanket rollout. This double role (consensus-roadmap and cautionary example) is noted in recent overviews. Radiology's earlier experimentation reveals the promise of AI-enhanced education to open access to diverse cases, to free up faculty times, and to encourage precision learning. But these very same trials illuminate the hazards of over-reliance, of misinformation, of biased output, and these cannot be the byproducts of afterthoughts but require careful pedagogical design. As a test case of AI-enhanced education, radiology holds implications not only for itself but, by extension, for the broader medical educational landscape, within which comparable dynamics will inevitably arise with the expansion of digital infrastructures.

3.2. Precision and Personalization in Radiology Education

The idea that radiology education could move beyond standardized curricula to individualizable pathways gained new life with the addition of artificial intelligence. The notion of "recision medical education" came to life first with Duong et al. (2019), who drew an analogy to precision medicine: just as therapy becomes individualized to patients' molecular and clinical profiles, so educational intervention could be individualized to the cognitive profiles, skill sets, and learning patterns of individual learners. Radiology, with its workflow-based education and case-based reasoning, is fertile ground particularly suited to this analog. Evidence to date backing this direction has grown exponentially. Hui et al. (2025), in a scoping review of generative AI's uses to date in radiology education, uncovered multiple platforms that personalize learning experiences by dynamically adjusting case difficulty, by modeling conditions that will only arise infrequently, and by offering automated testing commensurate with individual competence. Such systems reduce over-reliance on serendipitous exposure to cases, perennial issues with radiology training whereby rotations will inexorably over-represent common findings and pass over rarer conditions of crucial importance to diagnosis. By systematic correction of these gaps, AI-driven curricula eliminate case-based disparities and standardize diagnostic readiness across diverse training sites.

Personalization here means more than mere case presentation. It includes adaptive sequencing, whereby delivery of the learning material is determined by a learner's

dynamic knowledge state. A student perpetually misdiagnosing postoperative changes as recurrent tumor, by way of example, can be guided through a thoughtfully developed sequence of opposing cases with explanatory feedback. This is aligned with deliberate practice guidelines where repeated experience of judiciously tuned difficulties with immediate corrective feedback results in durable skill attainment. Moreover, by assessing longitudinal performance data, AI systems can establish individualized learner profiles that pinpoint obstinate Weaknesses, solidify notional Strengths, and provide instructor-led remediation guidance.

Another aspect of personalization is feedback loops. Computer-based performance monitoring can provide learners not only with knowledge of correctness of interpretation, but with understanding of why specific mistakes were made, promoting metacognitive understanding. Such systems prompt students to practice self-regulated learning, pondering their process of reasoning and proactively adjusting strategies. This resembles known theories of cognitive apprenticeship, whereby beginners internalize expert reasoning through explaining justifications and exercising scaffolded guidance. In computer-mediated learning, synthesized reports or diagnostic explanations emerge as subjects of reflection instead of terminal points, promoting reflective practice being core to building independent expertise.

Personalization is of special potential to offset geographic and institutional disparities. In low-resource or high-volume training settings, exposure to diverse case material and repeated faculty attention is not standard. Personalized learning with artificial intelligence can provide a kind of compensatory equity by making sure that students in these settings nevertheless get repeated exposures to diverse diagnostically rich material and tailored feedback. By no means does this eliminate mentorship or immersion inequalities, but to a lesser extent, educational outcomes depend on institutional resources.

These advances, nevertheless, carry their costs. The adaptive platforms can unintentionally reinforce then-prevailing biased results should their training sets undersample individual pathologies, populations, or image modalities. Automation-based personalization can also limit rather than augment learners' repertoires of cognition, producing algorithmic-hint dependences rather than diagnostic autonomy. The solution then lies in integrating the personalization within a supervised, human-in-the-loop system by which faculty guarantees that the adaptive resources augment rather than supplant deliberate reasoning.

As a whole, the shift to precision education ranks among the largest pedagogical advances in radiology education. By freeing oneself from efficiency to learner-centric design, technologies of artificial intelligence promise to tailor curricula not only to cohorts but to individual trajectories. The potential gain is vast: improved diagnostic preparation, institutionally consistent exposure, and more equitable training outcomes. But to reap such advantages, personalization will need to be undertaken with transparency, with

scrutiny, and with careful attention to the enduring goals of medical education: autonomy, critical thinking, and professional discernment.

3.3. Engagement and Simulation

3.3.1. Gamification, Simulation, and Engagement

Radiology education has long been panned for its heavy usage of didactic lectures and fixed case reviews, formats that frequently cannot maintain learner interest or induce active involvement. The addition of AI-powered gamification and adaptive simulation has thus been received with enthusiasm as enhancing learning from a passive to an iterative, stimulating, and feedback-rich experience. Such innovations represent not only technological improvements but pedagogical treatments that closely align with accepted learning and motivational theories. Radiology gamification refers to the integration of game elements, such as competition, achievement badges, and scoring that occurs in real-time, into diagnostic training modules. When combined with adaptive software algorithms, gamified platforms can dynamically alter case difficulty, keeping the challenge within an optimal difficulty domain. This directly borrows from the principle of self-determination theory, whereby learner motivation is maintained by feelings of competence, autonomy, and relatedness (Ryan & Deci, 2000). Through correct answers being reinforced while error patterns are fed back to learners constructively, AI-powered gamification establishes a reinforcing feedback loop that stimulates repeated usage and deliberate practice.

The evolution of RadGame is an excellent example of this model. Baharoon et al. (2025) found, in one recent review, that students using the platform showed significantly greater gains in lesion localization and report-writing proficiency than controls trained using traditional, passive approaches. The platform exploits adaptive algorithms that align case difficulty with learner proficiency, presenting increasingly difficult challenges that mirror the dynamic learning curves of true clinical practice. Such systems differ in important ways from preceeding efforts to create educational games by embedding diagnostic reasoning directly within gameplay, with the exercise being both fun and constructively educational. Simulation is a separate but complementary area of innovation. Adaptive AI-based simulations offer simulated environments within which trainees can practice interpretive procedures, navigate diagnostic uncertainty, and receive formal feedback without danger to patients. Negrete et al. (2025) detail the application of such systems in dentomaxillofacial radiology, where automated Objective Structured Clinical Examination (OSCE) testing and decision-support modules present learners with repeated practice of realistic scenarios. By building AI into simulation, teachers can create rare or high-stakes cases that cannot be accessed during clinical rotations and thus enhance exposure and diminish educational inequalities.

The educational value of simulation lies in its alignment with Kolb's (1984) model of experiential learning, through which experience's transformation to learning is effected by the sequence of concrete experience, reflective obser-

vation, abstract conceptualization, and active experimentation. Simulation with AI expands this sequence by providing immediate personalized feedback, shortening the span between the execution of a task and reflection. The learner not only can receive the outcome of their scoring but is prompted to review their rationale, assess the recommendations made by AI, and practice cases with conditions varied. Repeating this process strengthens cognitive flexibility and develops broader diagnostic understanding.

Also relevant here is the connection between gamification, simulation, and cognitive load. Radiology education frequently requires heavy perceptual and cognitive processing, potentially overwhelming new learners. Through organization of practice by adaptive difficulty and step-wise challenge, AI-mediated learning environments regulate cognitive load, keeping learners within the optimal space between boredom and overload. The act of scaffolding thus ensures that mastery is not attained through sacrifice of understanding, but aids the slow internalization of the heavy diagnostic schemas.

In tandem, gamification and simulation underscore how radiology training can fundamentally redefine its affective and cognitive elements. They foster repeated practice, maintain motivational traction, and enhance exposure to varied cases, with all of this being done through the integration of adaptive feedback mechanisms that solidify learning. But the potential of these innovations relies on thoughtful curricular integration. Left unchecked, gamified sites could afford too much emphasis on fun rather than rigor, or simulations might engender overdependence on algorithmic feedback rather than self-reliant reasoning. For that reason, each of these tools best works not through substitution of traditional pedagogy, but rather through augmentative strategies that can help to reinvigorate radiology education through better alignment of practice with the psychological and cognitive truth of how expertise emerges.

3.3.2. Teaching, Learning, and Engagement Innovations

Traditional radiology education has been based on the integration of didactic lectures, fixed case reviews, and hands-on experience during clinical rotations. Effective in the delivery of content, they frequently falter in maintaining learner interest or responding to individual learners' divergent needs. The entry of large language models (LLMs) opens new horizons in rethinking instructional design, allowing instructors to formulate lesson planning, tailor the content to learners of diverse knowledge levels, and innovate interactive pedagogies that actively engage learners within the learning process. Abd-alrazaq et al. (2023) observe that educators can ask LLMs to prepare lesson outlines, case-based activities, or different descriptions of abstract concepts, thus alleviating preparation hours. Jowsey et al. (2023) go a step further to show that while outputs generated by AI itself will seldom be ready to be inserted into classrooms, they yield fertile drafts to be edited by faculty. It streamlines planning and guarantees that curricular content is kept sensitive to learners' needs. Notably, this workflow repositions the teacher's role from iterative

drafting to higher-order functions of curation, contextualization, and mentoring.

Engagement is not about efficiency alone; learning itself is integral to engagement. Theories of cognitive engagement stress that the active involvement of learners and feedback-rich settings are imperative to knowledge retention (Prince, 2004). LLMs can be utilized to create interactive learning experiences like case challenges, quizzes, and problem-based discussion that maintain learner enthusiasm. Kasneci et al. (2023) demonstrate how prompts created by AI can refresh old content by delivering its new formats, ranging from gamified quiz to diagnostic puzzles. By integrating interactivity into the design of lessons, teacher-educators can transcend receptive passivity and encourage richer cognitive processing.

Another important area is visual learning. Pictures that can be utilized to visualize scientifically rich anatomical or pathological concepts can be created with the help of text-image models such as DALL-E 3, Stable Diffusion, and Midjourney (Artsi et al., 2024). Pictures that reduce cognitive load by providing dual coding (verbal and visual representation) in tandem can help learners learn difficult material (Mayer, 2009). However, such educators will need to understand the limitations of the current-day text-image tools to produce radiologically correct images, particularly with novel pathologies. The pictures will be susceptible to confusing learners or diluting diagnostic sophistication without strict expert control.

Beyond static resources, LLMs are being increasingly utilized to provide scaffolding of clinical reasoning. Research indicates that differential diagnoses made by LLMs correspond with expert opinion in 63–69% of cases given text-based descriptions of imaging (Kottlors et al., 2023; Sarangi et al., 2024). Such tools enable learners to experiment with structured processes of reasoning, test hypotheses, and access contextual explanation of image findings. For instance, a question such as, “Explain why miscarriage is more likely in septate uterus versus uterus didelphys,” can trigger rich explanation that clarifies not only outcomes but the very pathophysiology (Jowsey et al., 2023). Utilized thus, LLMs offer a kind of scaffolding that augments instruction by the faculty while stimulating learners to interrogate their reasoning actively.

The pedagogical promise carries over to workplace learning. Once AI tools become increasingly embedded within clinical practice, they can potentially identify mismatches between trainee and attending interpretations, select individualized teaching cases, and suggest practice tasks with specific focus. This resembles the model of precise medical education, whereby training is adjusted and aligned with the learner's progression dynamically. Such integration will potentially turn workplace supervision into ongoing, adaptive mentorship with real-time feedback. Even with these opportunities, however, considerable obstacles remain. Too much teacher content produced by artificial intelligence will reduce pedagogical diversity, and unchecked usage of explanation-based artificial intelligence will reinforce superficial knowledge. Professors will further have to learn to critically evaluate artificial intelligence outputs prior to

their usage to teach, lest educational rigor falls prey to ease. The challenge here is therefore that of integration: employing artificial intelligence to enrich engagement while securing the interpretive control of professors and the students' critical thinking.

3.4. LLMs in Radiology Education

3.4.1. Opportunities and Risks of LLMs in Radiology Education

The sudden arrival of large language models (LLMs) like ChatGPT has generated significant interest in their educational potential, and radiology was one of the first specialties to adopt such technology. Unlike their predecessors of decision-support software, LLMs were created with naturalistic interaction in mind, allowing students to have tutoring-simulation conversational exchanges with the technology. They can explain intricately technical terms, create study guides, and offer rapid feedback, which makes them general-purpose knowledge-acquisition and skill-reinforcement aids.

Other research records the educational value of LLMs as experienced by students. Students note that they can access these resources to learn controversial topics, craft practice questions, and access iterative, need-based feedback (Abouammoh et al., 2023; Alyasiri et al., 2024; Hosseini et al., 2023; Temsah et al., 2023). Interactive functionality aids student-driven learning and instills autonomy, desirable attributes particularly within radiology training and diagnostic independence being an eventual outcome. Accounts also report students practicing within resource-scarce settings valuing LLMs as cost-effective additions to texts or traditional tutorials, assertive regarding the contribution of these resources within expanding educational equity (Busch et al., 2023; Friederichs et al., 2023; Giannos, 2023; Gilson et al., 2023; Grabb, 2023; Guo & Li, 2023; Sedaghat, 2023).

The applications of LLMs extend from being cost-effective to scalable. They can provide conversational learning support at any time of the day without the constraint of geographical locations. In contrast to teachers who work within the bounds of time and location constraints, teachers can provide continuous learning support at any time of the day and can adapt their explanations according to the learner's prior knowledge (Abd-alrazaq et al., 2023; Kasneci et al., 2023; Yan et al., 2024). The ability of teachers and learning applications to provide continuous learning support according to the learner's needs fits the concept of active learning models in which practice and goal achievement through constant learning support from teachers can be improved through proper adaptive learning support and goal achievement (Ericsson, 2008; Prince, 2004; Ryan & Deci, 2000; van der Vleuten et al., 2012). In radiology education, learning opportunities are often constrained by limited access to diverse case materials and imaging examples. This limitation can be addressed through AI-enabled learning applications that extend teacher support by providing continuous, adaptive learning experiences and individualized feedback, thereby expanding trainees' exposure to var-

ied radiological cases (Artsi et al., 2024; Duong et al., 2019; Mistry et al., 2024; Tejani et al., 2022).

Yet the very qualities that make LLMs desirable carry with them enormous peril. Stochastically generated, not all outputs will be accurate, and the tendency to construct plausible but incorrect answers - "hallucinations" - pose significant danger to new learners. Tangadulrat et al. (2023) comment on this gap between perception: whereas 83% of medical students rated ChatGPT outputs appropriate learning material, only 53% of attending physicians agreed, reflecting faculty concerns about correctness, richness, and trustworthiness. Herein lies a bigger pedagogic concern: enthusiasm among students will not always translate into defensible pedagogy without regulation.

Ethical questions contribute to integration difficulty. Plagiarism, attribution, and academic integrity questions have been raised (D'Antonoli et al., 2024; Grunhut et al., 2021; Meyer et al., 2023; Pal et al., 2024; Weidener & Fischer, 2023). LLMs can craft eloquent paragraphs that students may be tempted to present as their own output, blurring lines between assistance and authorship. Without guidelines, LLM utilization threatens to erode core medical education principles of honesty and professionalism. Moreover, because many of these systems too often lack clarity about their data sources, students cannot safely distinguish between evidence-based knowledge and randomly generated knowledge, and spreading misinformation and trust erosion questions arise.

Cost is another component of risk. Whilst many LLMs now exist to be used gratis, sustaining access to high-calibre, medically specialized models will likely require subscriptions. Cascella et al. (2024) and Shah et al. (2023) note that such costs may place additional burdens on trainees already facing significant financial pressures, including the anticipated \$4,000 cost incurred on the pre-exam (Bhatnagar et al., 2019). If education through the use of AI becomes conditional on payment of subscriptions, rather than reducing inequalities of access, they may grow and hence widen inequities within the profession.

Transparency and regulation are perennial gaps. Commercial LLMs exempt themselves from the regulation that applies to medical software, and therefore students and teachers have to confront issues of bias, privacy, and data protection without institutional guidance (Blumenthal-Barby, 2023; Council, 2017; Hamed et al., 2023; Nashwan et al., 2023; Sheth et al., 2024; US Food and Drug Administration, 2024). It has already been determined that these models inherit and perpetuate their training data's biases, reinforcing stereotypes and injustices (Babool et al., 2022; Chen et al., 2021; Doo & McGinty, 2023; Louie & Wilkes, 2018; Parker et al., 2017; Poole-Dayana et al., 2024; Raj et al., 2024; Singh et al., 2023; Trabilsy et al., 2023; Zhu et al., 2024). For radiology education, with diagnostic accuracy and fair explanation being of greatest significance, these cannot be taken lightly. Considered together, LLMs' potential and danger in radiology education represent the general trade-off between innovation and prudence. To the benefit of wider accessibility, individualized learning, and scalable correction, these technologies can lead learners

astray, perpetuate disparities, and erode academic integrity. The dilemma of radiology's educators is not to adopt or not to adopt LLMs, but to adopt responsibly, via vetting, clarity, faculty stewardship, and ethical protection. The radiologic experience with LLMs will, by extension, provide models or foreshadowing for other medical specialties, and define both promise and conditions of success of conversational artificial intelligence as a legitimate education partner of precision.

3.4.2. Curriculum Development with LLMs

The creation and maintenance of superb radiology curricula were always manpower-intensive endeavors; ongoing updating is required to remain current with new advances in the imaging modalities, new diagnostic guidelines, and evolving accreditation requirements. By convention, this effort falls heavily on faculty with responsibilities to plan curriculum, practice clinically, conduct research, and teach. Large language models (LLMs) hold out new prospects by their ability to provide quick help with design of lessons, alignment with competencies, and iterative review of curriculum. Although LLMs by themselves will not obviate the necessity of expert advice by faculty, they hold promise of turning the process of curriculum building into a more dynamic and learner-focused process. Abd-alrazaq et al. (2023) show that LLMs can aid the entire curricular continuum, from creating course outlines to recommending assessment rubrics. Lee (2024) shows their benefit in streamlining course planning by studying learner performance data and recommending salutary intervention. Such uses demonstrate the potential of AI to decrease administrative burden while concurrently personalizing curricula to changing learner need. The outcome is a process of curriculum that is faster and more responsive, of particular value in radiology where advances in technology exceed conventional intervals of review of curriculum.

The incorporation of established educational structures reinforces the position of LLMs within curriculum design even further. Bloom's Taxonomy (Bloom et al., 1956; Conklin, 2005) is still seminal to expressing cognitive aims, while the Accreditation Council for Graduate Medical Education (Accreditation Council for Graduate Medical Education, 2019) milestones contribute speciality-specific standards. Sridhar et al. (2023) demonstrate that with careful prompting, GPT-4 can produce learning objectives aligned with these structures using the correct action verbs and cognitive levels. It illustrates the potential for including LLMs within formalized curriculum design, not to provide novelty but to help with alignment and coherence.

LLMs can provide capacity of formative curriculum review; faculty can input existing syllabi or schedules of rotations and ask the model to make recommendations about revisions, align material with milestones, or find gaps in coverage. The result is an iterative feedback loop that enables ongoing curriculum improvement without the delay of formal committees. Beyond objectives, LLMs can produce teaching scenarios, clinical cases, and practice questions that embed theoretical material within realistic diagnostic problems. A faculty member might ask, by way of example:

"Develop a teaching case demonstrating ultrasound artifacts appropriate to first-year residents, aligned with ACGME physics milestones." The resulting scenario serves as a starting point that can be refined by faculty with attention to clinical detail and educational sophistication rather than origination. The ability of LLMs is further extended by being integrated with other artificially intelligent systems. Programs such as Explainpaper, SummarAI, and Elicit have been applied to abstract journal articles, facilitate readings, and prepare students' self-testing questions (Kung et al., 2023; Rathinasabapathy et al., 2023). Their integration into radiology curricula can provide students with adaptive learning resources tuned to their ability levels, and instructors with ease of resource selection and maintenance. This is part of a broader pedagogical trend toward adaptive instructional design, by which material is continuously adjusted to learner advancement and situational demands (Merrill, 2012).

Even with these benefits, critical scrutiny is necessary. Scenarios or objectives generated by LLMs can seem reasonable but have inherent errors or mismatches with accreditation standards. The expertise of faculty is still necessary to maintain accuracy, contextual relevance, and educational integrity. Additionally, curriculum design based on AI raises new ethical and governance concerns. If curricula are developed based on models learned through transparent data, concern exists that internalized bias might affect what is learned and prioritized with adverse potential to perpetuate inequalities of diagnostic training.

In this context, LLMs will best fit into curricula as collaborators, not authors. They can speed up drafting, help align with educational frameworks, and provide wider repertoires of accessible teaching cases. But ultimately, responsibility will have to remain with faculty, whose clinician's eye and pedagogical experience can provide needed protection. Used responsibly, LLMs can make radiology curricula both more adaptive and rigorous, shortening the gap between fast-paced technological advances in practice and slow curricular reform cycles.

3.5. Learner Assessment with LLMs

Assessment has been a staple of radiology education, influencing not only learner advancement but programmatic accountability. Commonly, radiology assessment is characterized by multiple choice questions (MCQs), oral exams, and guided report writing. Each of these approaches possesses advantages and disadvantages: MCQs afford scalability but frequently lack realism; orals assess reasoning but require significant resources; report assessments mimic genuine practice to some extent but experience variability of grade. The arrival of large language models (LLMs) ushers in new possibilities to supplement these methods, potentially to make assessment truly continuous, scalable, and connected to genuine professional activity.

Radiology report writing is perhaps one of the most immediate areas of influence. At the draft stage, LLMs can abstract results, recommend structural organization, and make stylistic adjustments. Kasneci et al. (2023) comment that these capabilities, deployed longitudinally, enable in-

structors to monitor the progression of a learner's diagnostic reasoning and communicating over time. Instead of episodic assessments that are tied to individual rotations, feedback assisted by AI can document a continuous track of improvement that identifies patterns of diagnostic correctness, linguistic precision, and depth of interpretation. Here, LLMs become motors of formative assessment, offering immediate feedback while always deferring final decision to faculty.

Even summative testing is evolving. Olney (2023) piloted questions developed by Bing Chat and Macaw in anatomy and physiology and determined that human raters scored them to be between 91–93% informative as human-constructed questions. Error rates were not nil; Macaw sometimes repeated its distractors (7.5% of cases) and Bing Chat omitted correct answers in 9%. However, the findings indicate LLMs can be utilized effectively to contribute to item banks. Given that creating plausible distractors is among the most labor-intensive processes of test building (Gierl et al., 2017), freeing up faculty-time through AI-assisted item building is promising and will probably necessitate expert approval even after freeing up the time. To these overall conclusions, a study specific to radiology determined that GPT-4 created board-style questions and rationales of ACR DXIT exam item quality with perfect correctness of answers (100%), demonstrating its potential to contribute to exam preparation with an emphasis on specialties (Mistry et al., 2024).

Theory of assessment emphasizes the necessity of striking a balance between efficiency and validity and reliability. On Messick's unified account (Messick, 1995), validity is not inherent in the test itself but rather of interpretations and uses that are made of scores. In radiology, then, this implies that MCQs or report critiques produced by AI cannot be evaluated by face plausibility alone; they will have to reliably embody defensible inferences about competence. LLMs should thus be framed not as coming to replace validated instruments but rather to populate test pools to a wider extent, to help scaffold formative feedback, and to pinpoint areas of concentrated faculty assessment.

Yet another new frontier is that of programmatic measurement, with summation of multiple low-stakes measures over the long term to make higher-stakes judgments (van der Vleuten et al., 2012). LLM-driven feedback of any kind (report writing, case reasoning, or self-evaluations) can provide a stream of micro-measures. When systemically aggregated, these data can yield rich longitudinal profiles of learner progress and complement traditional summative tests with finer profiles of competence and broader item banks within specialty-specific arenas like radiology (Mistry et al., 2024).

Nevertheless, risks abound. Over-reliance on questions generated by machines may reduce diversity and inadvertently reduce the scope of measurement. Students may also learn to imitate stylistic patterns of pieces generated by machines, compromising the integrity of measurement. Even more fundamentally, there exists the danger of preferring that which can easily be generated by machines (schematized texts, MCQs) to that more hard to capture

(clinical judgments, interpersonal competence). There is an echo here of perennial criticisms of testing systems that reward that which can be measured to the expense of that which is valuable. To forestall these dangers, AI-driven tests will need to remain firm within faculty-led control. Human scrutiny cannot be avoided to ensure clinical validity, equity, and congruence with desired outcomes. Moreover, questions generated by machines should be tempered with performance-based exams such as OSCEs, workplace-based exams, and structured oral exams. Thus, LLMs serve to accelerate the rate of innovative testing rather than replace that which already exists. Taken together, LLM inclusion in radiology testing holds promise and peril. Their greatest promise lies within constructivist learning environments (continuous feedback, practice question generation, and learner self-assessment supporting scaffolds) where their use in summative measurement within high-stakes testing demands careful handling. By positioning these resources within a programmatic testing system, instructors can capitalize on their efficiency without compromising rigor, validity, or developing independent professional judgment.

3.6. Curricular Integration and Faculty Readiness

Whereas technological potential has been extensively proved, the hard part is integrating AI into the formal curricula of radiology training programs. The literature continuously reveals a gap between realization of AI's significance and its systematic integration into educational frameworks. Surveys and reviews indicate that whereas educators identify the unavoidability of AI in radiology, institutional capacity to facilitate curricular reorganization is limited. The discrepancy reflects a key paradox: enthusiasm about new technologies is strong, but willingness to ground and sustain it within organized forms is uneven. Doherty et al. (2025), in their European image educators' survey, found robust acknowledgment of relevance of training across AI but negligible trust in their capacity to provide effective delivery. Factors raised by the respondents included institutional inertia, scarcity of training resources, and lack of agreed-on frameworks as perennial obstacles. The observations correspond to wider global patterns. Huisman et al. (2021) demonstrated that whereas close to four-fifths of the respondents found inclusion of AI in residency curricula desirable, they concurrently raised doubts about inadequate knowledge among staff and unresolved ethical and legal issues. The similarity of these results across regions indicates that the challenge is not lack of intentions but lack of infrastructure and preparation among instructors.

The requirement for longitudinal planning has been emphasized by multiple authors. Gorospe-Sarasúa et al. (2022) and Gowda et al. (2022) recommend teaching of AI concepts even while in medical school, with reinforcement during residency and continuation through continuing medical education (CME). Tajmir and Alkasab (2018) and Tejani et al. (2022) make corresponding recommendations of vertical integration through the training continuum, with learners being exposed to AI not as an elective of late training but instead as an iterative theme aligned with clinical mile-

stones. All of these opinions correspond to general theories of curricular adjustment, which stress the importance of offering continuity and reinforcement to allow new competency to take hold. Literacy regarding AI cannot be added to an already-in-place curriculum by virtue of an independent module; instead, it must be incorporated through the learning process with supporting scaffolds that evolve with learners' increasing clinical responsibilities.

Faculty preparation is the key to this transformation. Unless teachers themselves are comfortable with the technical and pedagogical elements of AI, integration will struggle to move beyond superficial tokenism. Surveys bear out that many faculty staffs are not yet comfortable teaching about AIs, and that this is not due to lack of interest but unfamiliarity. This aligns with Rogers' seminal theory of diffusion of innovations (Rogers, 2003), whereby opinion leaders and early adopters play a crucial role in legitimating change. Early adopters of AI integration within radiology can thus provide examples of effective practice, mentor others, and make complexity seem less formidable. Institutional investment therefore is necessary to provide structured faculty education, in the form of workshops, certification schemes, and collaborative models of teaching. There are organizational issues too. Faculty report short of time, competing pressures, and lack of reward as excuses for not getting to grips with AIs education. The literature on change management explains that institutional backing, adjustment of workload, and reward mechanisms represent key levers of adoption (Kotter, 1996). Unless staff are both resourced and incentivized to adopt meaningful integration of AIs into their teaching, initiatives will lack coherence and will not be sustainable.

Finally, the literature draws a common picture: radiology trainees are primed for AI, but faculties and curricula lag. A dual-step approach is necessary to bridge this gap of readiness. Institutions need to invest in training educators with the knowledge and pedagogical approaches to teach responsibly about AI. Curricular redesign then needs to be deliberate and long-term, integrating concepts of AI at various levels of training. Unless these steps are undertaken, the promise of education enhanced by AI will be kept, only that education will continue to graduate learners technologically savvy, but pedagogically underserved.

3.7. Professional Perceptions and Generational Divides

Opinions about artificial intelligence in radiology exist widely, and the disparity between professional organizations, generations, and geographic locations directly influences how education integrates this technology. Surveys consistently report that radiology ranks among specialties that will be significantly impacted by AI, and the degree of enthusiasm, trust, and concern considerably differ among respondents. The differences not only establish curricular emphases but also the overall professional atmosphere within which radiology residents are socialized.

Naeem et al. (2025), during a national questionnaire of physicians across Pakistan, determined that a third of those questioned pinpointed radiology as the most affected by

AI. Notably, those questioned were in favor of including AI within medical curricula, referencing the specialty's need for rigorous programmed workflows and image interpretation. Such feedback represents an acknowledgement of AI's inevitability but rather betrays a pragmatic concern: without appropriate training, radiology is exposed to disruption rather than enhancement by tech revolution. Endorsement of education about such technologies represents less optimism than professional imperative. Within the United States, Alarifi (2025) recorded a comparable optimism among younger radiologists, recording higher familiarity with the tools of AI and stronger expectations of benefit. The same cohort, however, showed higher vulnerability to errors faced with erroneous outputs of an AI, suggesting a paradox of early adopter: while eager to exploit new technologies, such individuals can be equally susceptible to their shortcomings. Such contradiction is pertinent to technology adoption literature, reporting that early adopters will be tolerant of hazards and unknowns not borne by their older colleagues (Venkatesh et al., 2003).

Generational divisions within radiology transcend comfort with equipment to broader questions of professional identity. Younger trainees, who have grown up in a world surrounded by electronics, experience AI as an extension of their pedagogical world. Older radiologists, however, may experience AI as a danger to professional competence. The differing perceptions mirror broader patterns of professional identity formation, in which generations socialize into the profession with different assumptions about autonomy, expertise, and technology's role (Cruess et al., 2019). To younger trainees, AI may embody opportunity and relevance; to older radiologists, it may embody fears of erosion of interpretive judgment. Geographic contexts add further layers of complexity, such that in low- and middle-income countries (LMICs), enthusiasm about AI is tempered by resources, concern about expense, and lack of access to infrastructure. In high-income countries, resistance may be less about feasibility and rather about ingrained professional custom or uncertainty about medico-legal implications. Each establishes its own set of educational imperatives: in LMICs, education with AI may embody resource offset and access equity; in high-income countries, education about AI may embody a mechanism to ensure professional control within a rapidly automating world.

Educationally, these gaps are significant because they define legitimate knowledge. If students blindly accept outputs of AI while instructors are doubtful, the outcome can be a tacit curriculum that conveys ambivalence instead of clear guidance. The gap can only be closed by curricular measures that affirm enthusiasm and skepticism alike and position AI itself as a tool of which advantages and disadvantages depend on careful evaluation and judicious utilization. Instruction can be individualized by stage of career: younger trainees can be initiated to proper deference by guided scrutiny of interpretations generated by AI, while older instructors might require upskilling to teach with new pedagogic tools with confidence.

Finally, perceptions of radiology's future with AI depend on broader questions about the practice itself. Is radiology

characterized by human interpretive expertise that can only be supplanted by AI, or by a dynamic collaboration between man and machine? The response is generation-, geography-, and practice-specific. Curricula that don't recognize these divides will leave students trapped between competing narratives. Curricula that make these divergences explicit and that teach students to traverse them can turn potential fount of fragmentation into chance of intergenerational discussion and practice revitalisation.

3.8. Behavioral Risks and Human-in-the-Loop Safeguards

3.8.1. Behavioral Risks: Overreliance and Skill Retention

Whereas the inclusion of radiology education with AI holds many advantages, it concurrently poses behavioral hazards that can compromise long-term skill maintenance and professional autonomy. Primordial among these is over-reliance, whereby students rely too heavily on algorithmic interpretations, gradually diminishing their ability to make independent diagnostic judgments. This is not unique to radiology, but the subspecialty's heavy dependence on visual interpretation and templated reporting places it uniquely at risk. Sunday (2025) cautions that routine usage of AI can engender diagnostic deskilling, complacency, and attenuation of situational awareness. Such hazards multiply within educational settings, where learners may respond to AI reportage as definitive rather than as learning stimuli. Left unchecked, such over-reliance can engender a hidden curriculum whereby learners internalize passivity rather than critical review. Li and Little (2023) emphasize the necessity of promoting proper over-reliance, an architecture whereby AI serves as an aid to rather than substitute for human expertise. This necessitates teaching strategies that explicitly balance efficiency with goal-oriented practice to ensure that learners not only continue to hone visual acuity and reasoning abilities but actively employ such abilities while interacting with and using AI technologies.

Insights borrowed both from cognitive psychology and human-AI interaction research bear out these suspicions. Automation bias, the researched-and-confirmed tendency to over-rely on machine output even when incorrect, was found to be present within the context of clinical decision support (Goddard et al., 2012). In radiology, this could manifest as blind agreement with results highlighted by AI, reducing vigilance and impairing diagnostic reasoning. Cognitive offloading works to contribute to the peril: when students frequently offload interpretive endeavour to AI, they may save short-term cognitive resources but lose encoding of rich, structural knowledge that is key to expert performance (Carr, 2020).

Mitigation strategies are starting to appear in the literature. One is alternating between AI-assisted and AI-free case interpretation, with learners practicing both independent reasoning and augmentation of their workflow with AI. Another is including explainability prompts within AI platforms, with users being asked to explain their reasoning

before their output is compared with that of the AI. Such prompts turn the AI into a source of reflection rather than answers, which is consistent with deliberate practice models of expertise building. Assessment tasks can similarly be engineered to encourage independent interpretation and thus signal the ongoing supremacy of human judgment. The broader educational challenge is to reset the perception of AI not as displacing human cognition but rather to make us better engaged critics. Students need to learn to question AI output, not just what the outcome is but why it is, how well it adheres to established reasoning, and where possibly it is being misled. Unless we adopt this reflective stance, the danger is that radiology training will standardize complacency instead of building the expertise that defines the radiologist.

Hence, behavioral risks are not secondary to the inclusion of AI in education but foundational to its pedagogical design. By identifying overdependence and loss of skill as structural hazards, teachers can embed protection that ensures the integrity of training despite reaping the benefit of the effectiveness and flexibility of AI. The learning of radiology with these risks will likely have implications applicable to other specialties that are now set to embark on learning environments with aid by AI.

3.8.2. Human-in-the-Loop Safeguards

The dangers of over-reliance and deskilling posed above therefore require to be actively protected against to prevent the perpetuating of AI reinforcing instead of displacing human expertise. One closely common solution is the human-in-the-loop (HITL) model, which stresses that products of any AI will inevitably need to be interpreted, contextualised and verified by human end-users and not accepted on face value. In radiology training, HITL frameworks make AI an educational collaborator instead of an independent decision-maker whereby trainees actively contribute to the diagnostic process.

Oye et al. (2025) outline a formalized model of radiology education using HITL, based on interpretability, feedback loops, and systematic quality checks. By asking students to challenge outputs of AI (why a conclusion was reached and how that relates to their own line of reasoning) HITL systems foster critical rather than passive review. Such processes map onto reflective practice theories (Schön, 1983), such that learning occurs through iterative loops of action and reflection. Drafts or interpretations by AI then act as stimulants to criticism, with students prompted to identify error, hone their reasoning, and solidify their autonomy. Professional organizations have reasserted these agendas. A recent Association of Academic Radiology–Radiology Research Alliance task force white paper (2025) stressed that not a single system of AI should be implemented within clinical or educational settings without proper verification, scrutiny, and transparency (Ballard et al., 2025) Applying these mandates to education leads to that of AI tools that act rather than passively accepted supervisors. Thus, e.g., an AI-sourced report draft might act as the genesis of a teaching session during which trainees should be prompted to flag error, argue against alternate

interpretations, and resubmit conclusions. This converts that which might otherwise act as a danger of misinterpretation into an error-based learning exercise (Metcalf, 2017), with mistakes not avoided but actively utilized to gain understanding.

HITL safeguards also correspond to accepted practice of clinical supervision. As radiology attendings review trainee interpretations prior to signing off on reports, AI-assisted learning settings should integrate faculty review as a design requirement. Faculty thus fulfill a dual charge: they deliver corrective feedback and, no less importantly, embody proper skepticism of AI output. As experienced radiologists question and contextualize algorithmic recommendations, learners internalize not only interpretive competence but professional attitudes of prudence, responsibility, and skeptical thought. Another crucial aspect of HITL is that of transparency. Learners need to be made aware of the shortcomings of the AI system, its training data, error modes, and uncertainty contexts, and, unless this transparency is achieved, the learning process threatens to instantiate the very biases or blind spots that the AI system inherits with its datasets. Revealing these shortcomings within a supervision context enables learners to handle AI output responsibly rather than regarding them as dark oracles.

In sum, HITL safeguards reposition AI as a scaffold for critical engagement rather than a shortcut to answers. When implemented thoughtfully, they encourage learners to interrogate, critique, and improve upon AI-generated material, thereby strengthening diagnostic independence and professional judgment. Radiology's early experiments with HITL provide not only a model for other specialties but also a reminder that innovation without oversight risks eroding the very competencies education is meant to cultivate.

4. Recommendations for Future Development

The literature coalesces on the opinion that while radiology education's AI holds the potential to be transformative, its eventual worth will be less about the technology itself and instead about the system of governance, faculty preparation, and equity within which it is implemented. As the evidence matures, a set of priorities for planned development emerges. The recommendations go beyond cataloging advantages and disadvantages to sketching out circumstances within which AI can be a responsible collaborator with precision medical education.

4.1. Assessing Technological Readiness

The immediate concern is to have defensible indicators of when AI tools are mature enough to be included within educational environments. De (2023) stresses that putting individual learners or teachers on their own to judge model performance is unrealistic. The decision to be ready should rather be made through mutually agreed processes coordinated by professional associations. Yan et al. (2024) suggest, while scoping 118 articles, that innovations should be repeated using new models with new data, with open-source systems given higher priority to provide trans-

parency, and human-centered approaches to evaluation should be embedded. Such evaluations should go beyond accuracy and take into account explainability, equity, and transferability. A rigorous readiness system would make educators confident that adoption is not coming too soon while establishing common expectations across institutionally.

4.2. Faculty Development and Support

The success of any educational innovation ultimately rests with faculty. Even the best AI tools will not enhance learning without instructors feeling comfortable or competent to employ them effectively. Professional development initiatives requested by Akinici D'Antonoli et al. (2024) should prepare teachers to critically assess products of AI, make optimal use of prompting strategies, and responsibly embed tools into teaching practice. Such preparation should extend beyond technical proficiency to ethical issues of plagiarism, bias, and integrity. Faculty workshops, certification tracks, and common collections of best practice might provide the basis of sustainable building of capacity of faculty. Such initiatives align with wider scholarship on professional development of medical education, which stresses that adoption of innovation demands cultural and technical accommodation (Steinert et al., 2006).

4.3. Incorporating AI Training into Residency and CME

Integration of the curriculum should be longitudinal throughout undergraduate medical education into residency and continuing medical education (CME). The residency itself is another important watershed, too, with practicing physicians-in-training assuming practice without supervision. There should be structured modules on appropriate use of AI (based on clinical milestones) so that new graduates emerge with technical competence with ethical knowledge. Tajmir and Alkasab (2018) note that vertical alignment, with preliminary exposure to AI ideas early and repeated exposures throughout phases of training, is needed. The USACM (2017) principles, Hosanagar (2020), and the U.S. Government (2022) (n. p. House, 2022) provide points of comparison across curricular design. Through incorporation of the principles within formal education, teachers can foster a culture of reflective and ethical practice with regards to AI.

4.4. Governance and Equity

Governance mechanisms should be developed to protect against inequality and abuse. Considerations of cost, as Cascella et al. (2024) and Shah et al. (2023) have pointed out, open the door to the chance that access to superior AI resources might segregate learners by socioeconomic background. Equity thus necessitates deliberate policy, such as institutionally subsidized access or preference of open-source options. Data governance is similarly important. Misgivings over privacy, bias, and disinformation cannot be left to individual institutionally but need to be faced

through professional management and regulation. There may well be lessons here from proximate areas of digital health, where mechanisms of governance have developed to strike a balance between innovation and safeguarding patients (Topol, 2019).

4.5. Research Priorities

The evidence base is nascent, and research of a specific kind is needed to move beyond anecdote. Yan et al. (2024) identify some potential areas of focus: prevention of misinformation, resolution of inequalities, protection of privacy, and improvement of transparency. Beyond these, research is needed to assess long-term outcomes, such as the effect of AI-augmented training on maintenance of skills, professional formation of identity, and patient safety, including validity studies that compare radiology exam items generated by AI with known board materials (Mistry et al., 2024). Experimental designs of a rigorous quality, multi-institution studies, and mixed-methods studies will be valuable. Lacking empirical validity, adoption of the technology risks being driven by enthusiasm rather than evidence, a dynamic that undermines trust both within education and within clinical practice.

4.6. Toward a Balanced Future

These suggestions collectively demonstrate that the future of AI in radiology education is not predetermined but conditional. Tech potential must be aligned with readiness models, faculty training, integration into curricula, system of governance protection, and research. Radiology's leadership role in innovative solutions positions it to establish benchmarks that will have relevance throughout medical education. But with that capability comes a responsibility to not adopt too rapidly, to not adopt unevenly, and to not adopt without depth of ethical principle. Through planned, evidence-based approaches, the specialty can turn AI into a driver of precision medical education rather than disruption.

5. Discussion

The incorporation of artificial intelligence (AI) into radiology teaching represents a particularly informative case study with which to gain insight into opportunities and risks of educational innovation within medicine. Within the literature, a repeated theme is that whilst tools of artificial intelligence, such as large language models (LLMs) and generative platforms, can potentially revolutionize radiology trainee learning, the results depend significantly on the way that such tools are conceptualized, moderated, and integrated into established frameworks. The discussion assimilates the evidence borne, circle-backs to the research questions that guided the inquiry, and emphasizes wider implications within medical teaching.

Radiology is uniquely well situated to act as such a test case. Its organizational features, such as digital workflow, standardized reporting, and vibrant informatics culture, render it a perfect proving ground for educational uses

of AI. Systematic and umbrella reviews already have been shown to have such uses across report generation, case exposure, gamified learning, and adaptive simulation (Keshavarz et al., 2024; Lastrucci et al., 2024; Sacoransky et al., 2024; Temperley et al., 2024). Such uses document how AI can obviate historical difficulties, including uneven exposure to uncommon cases and delay of feedback, while concurrently revealing the hazards of early experience, including over-reliance, misinfo, and diagnostic deskilling. The dual role of radiology, therefore, both blueprint and cautionary tale, holds important lessons applicable to other specialties.

The mechanisms by which AI improves learning best can be grouped into three interconnected areas. First, adaptive case curation broadens exposure of learners to diagnostically informative patterns and gains entry to cases that would otherwise be too rare. Secondly, iterative feedback mechanisms implemented during practice transfer shift assessment to longitudinal rather than episodic frameworks and consequently enhance opportunities for self-regulated learning (Kasneci et al., 2023; Sridhar et al., 2023). Thirdly, language scaffolding explains to learners the process of report writing while, incidentally, exemplifying professional discourse and consequently facilitating technical and communication competence (Abd-alrazaq et al., 2023; Lee, 2024). All of these procedures align highly with already existing expertise development theories, such as deliberate practice (Ericsson, 2008), cognitive apprenticeship (Collins et al., 1991), and programmatic testing (van der Vleuten et al., 2012). Importantly, each reveals that AI not only can be a set of tools but can be regarded as a potential model of design of instruction that propels medical education toward greater precision.

Meanwhile, the literature highlights the peril of premature or uncritical adoption. Automation bias and over-reliance prove enduring issues (Goddard et al., 2012; Li & Little, 2023; Sunday, 2025), and the risk that students will accept AI output as definitive rather than to prompt reasoning. Ethics, including plagiarism, bias, and data management, add to this picture (Grunhut et al., 2021; Meyer et al., 2023; Pal et al., 2024). To respond to these fears, human-in-the-loop models have been set out, with a focus on interpretive control, error-based criticism, and reflection (Oye et al., 2025). Such systems, reflected in the Academic Radiology–RRA white paper (Ballard et al., 2025), turn AI itself into a learning scaffold rather than a source of risk. The model aligns with reflective practice theory (Schön, 1983) and error-based learning (Metcalf, 2017), positioning the AI not as oracle but rather as prompt to professional judgment.

In spite of mounting acknowledgment of the value of AI, adoption within the curriculum is uneven. Trainees are typically enthusiastic, yet surveys of staff reaffirm that staff will mostly not be confident that they can teach effectively about AI (Doherty et al., 2025; Huisman et al., 2021). The sort of asymmetry thus runs the risk of creating a generational gap in disposition to the embedding of AI (Alarifi, 2025; Naeem et al., 2025). Professional development then becomes a paramount imperative. The absence of repeated

investment in training teachers, institutional incentives, and mentorship by pioneers therefore causes integration of the AI to risk being tokenistic or piecemeal. The theory of the management of change (Kotter, 1996) and the model of the diffusion of innovations (Rogers, 2003) both argue that productive adoption will require not only individual enthusiasm but institutional backup and deliberate planning.

Even though much of this deliberation has centered on radiology, the implications go far beyond the specialty. Those specialties with similarly digitalized infrastructures, such as pathology, dermatology, and ophthalmology, will likely face equivalent opportunities and perils. By comparison, specialties with less standardized data will potentially face even higher levels of exclusion. What becomes apparent in each instance is that the fundamental concern is not adoption itself but design: incorporation of safeguards, reconciliation of educational usage with accreditation standards, and fair access across diverse learning settings. These developments, collectively, portend radiology's ability to help shape the future of precise medical education. By tailoring cases, feedback, and testing to individual learners, with the aid of AI, superior and more responsive educational systems can potentially coexist. Such potential, however, exists only conditionally. Unless proved through validation, transparent by design, and effectively governed, its potential can be derailed by the pitfalls of misinformation and lack of equity. The worry of educators, therefore, is to make AI's technical potential equivalent to pedagogical value, in effect, transmuting risk into possibility through reflective and thoughtful design.

6. Implications

The results of this research synthesis decisively affirm that the integration of artificial intelligence (AI) into radiology education cannot be conceived of as a simple technical issue of equipment adoption. Instead, it poses wider questions of pedagogy, professional formation, government, and equity. They carry implications that transcend radiology and determine the future direction of medical education across the board. The incorporation of artificial intelligence into medical education will have numerous implications for educators, curriculum developers, professional organizations, and research scientists. Educators will face the primary challenge of positioning AI not as a short-cut to increased efficiency but rather as a framework to afford deliberate practice. Effective teaching strategies will thus be those that require learners to challenge outputs of AI, question report generation, and shuttle between solo and AI-augmented activities. Curriculum developers will then intermix training with AI over an extended period of time, starting with basic exposure during medical school and continuing through residency and continuing professional education. Through this process, learners will understand the technology of AI and the judgment that their learners will need to employ the technology responsibly. The inclusion of constructive alignment principles—where learning objectives, activity, and assessing are made explicitly aligned (Biggs & Tang, 2011)—will be nec-

essary to ensure that tools enriched with AI attain desired outcomes and not provide portals to short-cuts.

Professional organizations play a key role in defining readiness standards, validation processes, and ethical guidelines for using AI in teaching. The Association of Academic Radiology–RRA task force assumed this role already, by providing direction through its white paper on large language models and radiology education (Ballard et al., 2025). Overall governance structures will need to tackle equity too, to prevent cost structures from stratifying access to AI-augmented learning. Here, examples of digital health governance (Topol, 2019) and of responsible innovation frameworks (Stahl et al., 2021) postulate how societies can balance between innovation and accountability. Their approval will probably frame the legitimacy of AI-enriched teaching agendas and their usage across various institutions.

The research agenda of radiology education and artificial intelligence (AI) is correspondingly very early. Whereas previous studies have evaluated short-term precision, ease of use, and learner attitudes, very little is known about longer-term skill maintenance outcomes, diagnostic autonomy, and professional identity. To address these questions, feasibility studies should then be followed by longitudinal designs, randomized trials, and mixed-method studies evaluating quantitative performance and qualitative learner experiences. Cross-institution collaborations will be necessary to make results generalizable across sites, rather than specific to single-site pilots. Notably, synchronizing future research with evidence-based medical education principles (Cook et al., 2008) will enhance the credibility of integrating AI and provide a robust foundation of evidence to construct policies.

Although radiology has been at the focal point of much of this discussion, its implications don't stop with this discipline. Digital-heavy specialties of pathology, dermatology, and ophthalmology enjoy the very same digital infrastructures and will confront very similar difficulties. Bolder, the radiology experience holds a broader principle applicable across medical education: technological developments will need to proceed alongside pedagogical planning, ethical protection, and preparation of faculty. So radiology's earlier exposure to artificial intelligence is both an exemplar and a cautionary note to other specialties now looking to the promise and danger of integration with artificial intelligence.

7. Conclusion

Radiology's experience with artificial intelligence (AI) illustrates both the promise and the peril of technological innovation in medical education. At the nexus of encoded data, standardized reporting, and computerized workflow, radiology has been an early testing ground for large language models (LLMs) and generative aids. The evidence synthesized here reveals that AI can widen access to exotic cases, speed feedback, and individualize learning, thus furthering the broader imperative of precision medical education. However, the same evidence reveals equally clearly that without careful safeguards, these advantages will be

undermined by overdependence, diagnostic deskilling, misinformation, and inequity. This review has made the case that the true pedagogical potential of AI rests not with automation per se, but with how it can structure critical thinking, reflection, and diagnostic autonomy. Human-in-the-loop processes, error-based learning design, and alternating between AI-augmented and human-only interpretation provide pragmatic ways of interacting with AI without compromising foundational abilities. Curricular adoption lags unevenly, frustrated by gaps in teacher preparation, institutionally motivated resistance, and generational fissures within professional dispositions. The future will depend on types of governance that specify standards of preparation, research agendas that assess long-term outcome, and professional associations that link validation and supervision.

The review translates to various design suggestions that can be applied to practice and future research:

1. Alternating reliance: Radiology trainees who alternate between AI-assisted and AI-free interpretation will demonstrate higher long-term diagnostic accuracy than those trained exclusively with AI support.
2. Prompting for critical reflection: The integration of rationale-critique and explainability prompts into AI outputs will enhance metacognitive activity and decrease automation bias relative to answer outputs.
3. Programmatic integration: The integration of low-stakes micro-assessments driven by artificial intelligence within programmatic models of assessment will provide superior indicators of learner progress than episodic high-stakes tests alone.
4. Faculty capacity-building: Faculty development programs that mix technical training with ethical frameworks will not only enhance faculty comfort with teaching AI but will also gain learner trust in AI-mediated curricula.
5. Protection of equity: The education centers that provide open or subsidised access to AI resources will reduce learning outcome inequality between socioeconomic classes compared to settings with individual subscription-based access.

These assertions underscore that the deployment of AI in learning is not a destination but a design problem—one that necessitates weighing creativity with accountability. Radiology, by virtue of being an early adopter, provides

both inspiration and caution: its achievements exemplify what can be accomplished, while its perils indicate what should be protected against. The wider message to medical education is that AI should not be framed as a substitute for expertise but rather a co-teacher of its nurturing. When pursued with clarity, control, and pedagogical stewardship, AI can augment the routes through which radiologists and, by extension, doctors across specialties gain independence, discernment, and preparation to practice within an AI-enabled future.

.....

Conflict of Interest / Competing Interests

The authors declare no competing or conflicting interests.

Acknowledgments

None.

Data Availability

No datasets were generated.

Ethics Approval and Consent

Not applicable.

AI Tool Use Disclosure

AI tool: ChatGPT. Provider: OpenAI. Version: GPT-5, Oct. 2025. Purpose: Language refinement. Verification: The authors reviewed and verified all output.

Preprint Disclosure

This article has not appeared as a preprint anywhere.

Third-Party Material Permissions

No third-party material requiring permission has been used.

Submitted: October 16, 2025 AST. Accepted: November 25, 2025 AST. Published: November 28, 2025 AST.
ISSN: 3068-773X



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Abd-alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, M., Latifi, S., Aziz, S., Damseh, R., Alabed Alrazak, S., & Sheikh, J. (2023). Large language models in medical education: Opportunities, challenges, and future directions. *JMIR Medical Education*, 9, e48291. <https://doi.org/10.2196/48291>
- Abouammoh, N., Alhasan, K., Raina, R., Malki, K. A., Aljamaan, F., Tamimi, I., ... Temsah, M. H. (2023). Exploring perceptions and experiences of ChatGPT in medical education: a qualitative study among medical college faculty and students in Saudi Arabia. *MedRxiv*, 2023–07. <https://doi.org/10.1101/2023.07.13.23292624>
- Accreditation Council for Graduate Medical Education. (2019). *ACGME data resource book: Academic year 2018–2019*. ACGME. https://www.acgme.org/globalassets/pfassets/publicationsbooks/2018-2019_acgme_databook_document.pdf
- Alarifi, M. (2025). Radiologists' Views on AI and the Future of Radiology: Insights from a US National Survey. *British Journal of Radiology*, tqaf222. <https://doi.org/10.1093/bjr/tqaf222>
- Alyasiri, O. M., Salman, A. M., Akhtom, D. A., & Salisu, S. (2024). ChatGPT revisited: Using ChatGPT-4 for finding references and editing language in medical scientific articles. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 125(5), 101842. <https://doi.org/10.1016/j.jormas.2024.101842>
- Artzi, Y., Sorin, V., Konen, E., Glicksberg, B. S., Nadkarni, G., & Klang, E. (2024). Large language models for generating medical examinations: systematic review. *BMC Medical Education*, 24(1), 354. <https://doi.org/10.1186/s12909-024-05239-y>
- Babool, S., Bhai, S. F., Sanderson, C., Salter, A., & Christopher-Stine, L. (2022). Racial disparities in skin tone representation of dermatomyositis rashes: a systematic review. *Rheumatology*, 61(6), 2255–2261. <https://doi.org/10.1093/rheumatology/keab809>
- Baharoon, M., Raissi, S., Jun, J. S., Heintz, T., Alabbad, M., Alburkani, A., ... Rajpurkar, P. (2025). *RadGame: An AI-Powered Platform for Radiology Education*. <https://doi.org/10.48550/arXiv.2509.13270>
- Ballard, D. H., Antigua-Made, A., Barre, E., Edney, E., Gordon, E. B., Kelahan, L., & Adams, S. J. (2025). Impact of ChatGPT and large language models on radiology education: association of academic radiology—radiology research alliance task force white paper. *Academic Radiology*, 32(5), 3039–3049. <https://doi.org/10.1016/j.acra.2024.10.023>
- Bhatnagar, V., Diaz, S. R., ... Bucur, P. A. (2019). The cost of board examination and preparation: An overlooked factor in medical student debt. *Cureus*, 11(3), e4270. <https://doi.org/10.7759/cureus.4168>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). McGraw-Hill/Open University Press.
- Bloom, B. S., Engelhart, M. D., ... Furst, E. J. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay.
- Blumenthal-Barby, J. (2023). An AI Bill of Rights: Implications for health care AI and machine learning—a bioethics lens. *The American Journal of Bioethics*, 23(1), 4–6. <https://doi.org/10.1080/15265161.2022.2135875>
- Busch, F., Adams, L. C., & Bressem, K. K. (2023). Biomedical ethical aspects towards the implementation of artificial intelligence in medical education. *Medical Science Educator*, 33, 1007–1012. <https://doi.org/10.1007/s40670-023-01815-x>
- Carr, N. (2020). *The shallows: What the Internet is doing to our brains*. WW Norton & Company.
- Cascella, M., Semeraro, F., ... Montomoli, J. (2024). The breakthrough of large language models release for medical applications: 1-year timeline and perspectives. *Journal of Medical Systems*, 48(1), 1–11. <https://doi.org/10.1007/s10916-024-02045-3>
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, 4(1), 123–144. <https://doi.org/10.1146/annurev-biodatasci-092820-114757>
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3), 6–11.
- Conklin, J. (2005). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives complete edition*. <https://www.jstor.org/stable/42926529>
- Cook, D. A., Bordage, G., & Schmidt, H. G. (2008). Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Medical Education*, 42(2), 128–133. <https://doi.org/10.1111/j.1365-2923.2007.02974.x>
- Council, A. U. P. P. (2017). Statement on algorithmic transparency and accountability. *Commun. ACM*.
- Cruess, S. R., Cruess, R. L., & Steinert, Y. (2019). Supporting the development of a professional identity: general principles. *Medical Teacher*, 41(6), 641–649. <https://doi.org/10.1080/0142159X.2018.1536260>
- D'Antonoli, T. A., Stanzione, A., Bluethgen, C., Vernuccio, F., Ugga, L., Klontzas, M. E., ... Koçak, B. (2024). Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagnostic and Interventional Radiology*, 30(2), 80. <https://doi.org/10.4274/dir.2023.232417>
- De, A. (2023). Statistical considerations and challenges for pivotal clinical studies of artificial intelligence medical tests for widespread use: opportunities for inter-disciplinary collaboration. *Statistics in Biopharmaceutical Research*, 15(3), 476–490. <https://doi.org/10.1080/19466315.2023.2169752>

- Doherty, G., Hughes, C., McConnell, J., Bond, R., McLaughlin, L., & McFadden, S. (2025). Perception of medical imaging educators on the addition of AI education to the medical imaging curriculum: A cross-sectional survey. *Radiography*, 31(6), 103153. <https://doi.org/10.1016/j.radi.2025.103153>
- Doo, F. X., & McGinty, G. B. (2023). Building diversity, equity, and inclusion within radiology artificial intelligence: representation matters, from data to the workforce. *Journal of the American College of Radiology*, 20(9), 852–856. <https://doi.org/10.1016/j.jacr.2023.06.014>
- 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. *The British Journal of Radiology*, 92(1103), 20190389. <https://doi.org/10.1259/bjr.20190389>
- Ericsson, A. K. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine*, 15(11), 988–994. <https://doi.org/10.1111/j.1553-2712.2008.00227.x>
- Friederichs, H., Friederichs, W. J., & März, M. (2023). ChatGPT in medical school: how successful is AI in progress testing? *Medical Education Online*, 28(1), 2220920. <https://doi.org/10.1080/10872981.2023.2220920>
- Giannos, P. (2023). Evaluating the limits of AI in medical specialisation: ChatGPT's performance on the UK Neurology Specialty Certificate Examination. *BMJ Neurology Open*, 5(1), e000451. <https://doi.org/10.1136/bmjno-2023-000451>
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, 87(6), 1082–1116. <https://doi.org/10.3102/0034654317726529>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312. <https://doi.org/10.2196/45312>
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>
- Gorospe-Sarasúa, L., Muñoz-Olmedo, J. M., Sendra-Portero, F., & de Luis-García, R. (2022). Challenges of radiology education in the era of artificial intelligence. *Radiología (English Edition)*, 64(1), 54–59. <https://doi.org/10.1016/j.rxeng.2020.10.012>
- Gowda, V., Jordan, S. G., & Awan, O. A. (2022). Artificial intelligence in radiology education: a longitudinal approach. *Academic Radiology*, 29(5), 788–790. <https://doi.org/10.1016/j.acra.2021.08.022>
- Grabb, D. (2023). ChatGPT in medical education: a paradigm shift or a dangerous tool? *Academic Psychiatry*, 47(4), 439–440. <https://doi.org/10.1007/s40596-023-01791-9>
- Grunhut, J., Wyatt, A. T., & Marques, O. (2021). Educating future physicians in artificial intelligence (AI): an integrative review and proposed changes. *Journal of Medical Education and Curricular Development*, 8, 23821205211036836. <https://doi.org/10.1177/23821205211036836>
- Guo, A. A., & Li, J. (2023). Harnessing the power of ChatGPT in medical education. *Medical Teacher*, 45(9), 1063–1063. <https://doi.org/10.1080/0142159X.2023.2198094>
- Hamed, E., Eid, A., & Alberly, M. (2023). Exploring ChatGPT's potential in facilitating adaptation of clinical guidelines: a case study of diabetic ketoacidosis guidelines. *Cureus*, 15(5). <https://doi.org/10.7759/cureus.38784>
- Hosanagar, K. (2020). *A human's guide to machine intelligence: how algorithms are shaping our lives and how we can stay in control*. Penguin.
- Hosseini, M., Gao, C. A., Liebovitz, D., Carvalho, A., Ahmad, F. S., Luo, Y., MacDonald, N., & Holmes, A. K. (2023). An exploratory survey about using ChatGPT in education, healthcare, and research. *PLOS ONE*, 18, e0292216. <https://doi.org/10.1371/journal.pone.0292216>
- House, W. (2022). *Blueprint for an ai bill of rights: Making automated systems work for the american people*. Nimble Books.
- Hui, M. L., Sacoransky, E., Chung, A., & Kwan, B. Y. (2025). Exploring the integration of artificial intelligence in radiology education: A scoping review. *Current Problems in Diagnostic Radiology*, 54(3), 332–338. <https://doi.org/10.1067/j.cpradiol.2024.10.012>
- Huisman, M., Ranschaert, E., Parker, W., Mastrodicasa, D., Koci, M., Pinto de Santos, D., & Willemink, M. J. (2021). An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education. *European Radiology*, 31(11), 8797–8806. <https://doi.org/10.1007/s00330-021-07782-4>
- Jowsey, T., Stokes-Parish, J., Singleton, R., & Todorovic, M. (2023). Medical education empowered by generative artificial intelligence large language models. *Trends in Molecular Medicine*, 29(12), 971–973. <https://doi.org/10.1016/j.molmed.2023.08.012>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Keshavarz, P., Bagherieh, S., Nabipoorashrafi, S. A., Chalian, H., Rahsepar, A. A., Kim, G. H. J., & Bedayat, A. (2024). ChatGPT in radiology: A systematic review of performance, pitfalls, and future perspectives. *Diagnostic and Interventional Imaging*, 105(7–8), 251–265. <https://doi.org/10.1016/j.diii.2024.04.003>

- Kotter, J. P. (1996). *Leading Change*. Harvard Business School Press.
- Kottlors, J., Bratke, G., Rauen, P., Kabbasch, C., Persigehl, T., Schlamann, M., & Lennartz, S. (2023). Feasibility of differential diagnosis based on imaging patterns using a large language model. *Radiology*, 308(1), e231167. <https://doi.org/10.1148/radiol.231167>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... Tseng, V. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Lastrucci, A., Wandael, Y., Barra, A., Ricci, R., Pirrera, A., Lepri, G., & Giansanti, D. (2024). Revolutionizing radiology with natural language processing and chatbot technologies: a narrative umbrella review on current trends and future directions. *Journal of Clinical Medicine*, 13(23), 7337. <https://doi.org/10.3390/jcm13237337>
- Lee, H. (2024). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*, 17(5), 926–931. <https://doi.org/10.1002/ase.2270>
- Li, M. D., & Little, B. P. (2023). Appropriate reliance on artificial intelligence in radiology education. *Journal of the American College of Radiology*, 20(11), 1126–1130. <https://doi.org/10.1016/j.jacr.2023.04.019>
- Louie, P., & Wilkes, R. (2018). Representations of race and skin tone in medical textbook imagery. *Social Science & Medicine*, 202, 38–42. <https://doi.org/10.1016/j.socscimed.2018.02.023>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Merrill, M. D. (2012). *First principles of instruction*. John Wiley & Sons.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741. <https://doi.org/10.1037/0003-066X.50.9.741>
- Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, 68(1), 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>
- Meyer, J. G., Urbanowicz, R. J., Martin, P. C., O'Connor, K., Li, R., Peng, P. C., & Moore, J. H. (2023). ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining*, 16(1), 20. <https://doi.org/10.1186/s13040-023-00339-9>
- Mistry, N. P., Saeed, H., Rafique, S., Le, T., Obaid, H., & Adams, S. J. (2024). Large language models as tools to generate radiology board-style multiple-choice questions. *Academic Radiology*, 31(9), 3872–3878. <https://doi.org/10.1016/j.acra.2024.06.046>
- Naeem, S., Abbas, H., & Hamna, S. (2025). Knowledge and perception of practicing doctors on artificial intelligence application in medicine. *JAIMC: Journal of Allama Iqbal Medical College*, 23(2). <https://doi.org/10.59058/97y7wk95>
- Nashwan, A. J., AbuJaber, A. A., & AbuJaber, A. (2023). Harnessing the power of large language models (LLMs) for electronic health records (EHRs) optimization. *Cureus*, 15(7). <https://doi.org/10.7759/cureus.42634>
- Negrete, D., Lopes, S. L. P. D. C., Barretto, M. D. D. A., Moura, N. B. D., Nahás, A. C. R., & Costa, A. L. F. (2025). Artificial Intelligence and Dentomaxillofacial Radiology Education: Innovations and Perspectives. *Dentistry Journal*, 13(6), 245. <https://doi.org/10.3390/dj13060245>
- Olney, A. M. (2023). Generating Multiple Choice Questions from a Textbook: LLMs Match Human Performance on Most Metrics. *Grantee Submission*.
- Pal, S., Bhattacharya, M., Islam, M. A., & Chakraborty, C. (2024). AI-enabled ChatGPT or LLM: a new algorithm is required for plagiarism-free scientific writing. *International Journal of Surgery*, 110(2), 1329–1330. <https://doi.org/10.1097/JS9.0000000000000939>
- Parker, R., Larkin, T., & Cockburn, J. (2017). A visual analysis of gender bias in contemporary anatomy textbooks. *Social Science & Medicine*, 180, 106–113. <https://doi.org/10.1016/j.socscimed.2017.03.032>
- Poole-Dayana, E., Roy, D., & Kabbara, J. (2024). Llm targeted underperformance disproportionately impacts vulnerable users. *arXiv Preprint arXiv:2406.17737*. <https://doi.org/10.48550/arXiv.2406.17737>
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223–231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- Raj, C., Mukherjee, A., Caliskan, A., Anastasopoulos, A., & Zhu, Z. (2024). Breaking bias, building bridges: Evaluation and mitigation of social biases in llms via contact hypothesis. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Vol. 7, pp. 1180–1189). <https://doi.org/10.1609/aies.v7i1.31715>
- Rathinasabapathy, G., Swetha, R., & Veeranjanyulu, K. (2023). Emerging artificial intelligence tools useful for researchers, scientists and librarians. *Indian Journal of Information Library & Society*, 36(3–4), 163–172.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sacoransky, E., Kwan, B. Y., & Soboleski, D. (2024). ChatGPT and assistive AI in structured radiology reporting: A systematic review. *Current Problems in Diagnostic Radiology*, 53(6), 728–737. <https://doi.org/10.1067/j.cpradiol.2024.07.007>
- Sarangi, P. K., Irodi, A., Panda, S., Nayak, D. S. K., & Mondal, H. (2024). Radiological differential diagnoses based on cardiovascular and thoracic imaging patterns: perspectives of four large language models. *Indian Journal of Radiology and Imaging*, 34(02), 269–275. <https://doi.org/10.1055/s-0043-1777289>

- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Sedaghat, S. (2023). Early applications of ChatGPT in medical practice, education and research. *Clinical Medicine*, 23(3), 278–279. <https://doi.org/10.7861/clinmed.2023-0078>
- Shah, N. H., Entwistle, D., & Pfeffer, M. A. (2023). Creation and adoption of large language models in medicine. *Jama*, 330(9), 866–869. <https://doi.org/10.1001/jama.2023.14217>
- Sheth, S., Baker, H. P., Prescher, H., & Strelzow, J. A. (2024). Ethical considerations of artificial intelligence in health care: examining the role of generative pretrained transformer-4. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, 32(5), 205–210. <https://doi.org/10.5435/JAAOS-D-23-00787>
- Singh, N., Lawrence, K., Richardson, S., & Mann, D. M. (2023). Centering health equity in large language model deployment. *PLOS Digital Health*, 2(10), e0000367. <https://doi.org/10.1371/journal.pdig.0000367>
- Sridhar, P., Doyle, A., Agarwal, A., Bogart, C., Savelka, J., & Sakr, M. (2023). Harnessing llms in curricular design: Using gpt-4 to support authoring of learning objectives. *arXiv Preprint arXiv:2306.17459*. <https://doi.org/10.48550/arXiv.2306.17459>
- Stahl, B. C., Akintoye, S., Bitsch, L., Bringedal, B., Eke, D., Farisco, M., ... Ulinicane, I. (2021). From Responsible Research and Innovation to responsibility by design. *Journal of Responsible Innovation*, 8(2), 175–198. <https://doi.org/10.1080/23299460.2021.1955613>
- Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher*, 28(6), 497–526. <https://doi.org/10.1080/01421590600902976>
- Sunday, O. (2025). Behavioral Impacts of AI Reliance in Diagnostics: Balancing Automation with Skill Retention. *Epidemiol Health Data Insights*, 1(3), ehdi011. <https://doi.org/10.63946/ehdi/16894>
- Tajmir, S. H., & Alkasab, T. K. (2018). Toward augmented radiologists: changes in radiology education in the era of machine learning and artificial intelligence. *Academic Radiology*, 25(6), 747–750. <https://doi.org/10.1016/j.acra.2018.03.007>
- Tangadulrat, P., Sono, S., & Tangtrakulwanich, B. (2023). Using ChatGPT for clinical practice and medical education: cross-sectional survey of medical students' and physicians' perceptions. *JMIR Medical Education*, 9(1), e50658. <https://doi.org/10.2196/50658>
- Tejani, A. S., Elhalawani, H., Moy, L., Kohli, M., & Kahn, C. E., Jr. (2022). Artificial intelligence and radiology education. *Radiology: Artificial Intelligence*, 5(1), e220084. <https://doi.org/10.1148/ryai.220084>
- Temperley, H. C., O'Sullivan, N. J., Mac Curtain, B. M., Corr, A., Meaney, J. F., Kelly, M. E., & Brennan, I. (2024). Current applications and future potential of Chat GPT in radiology: a systematic review. *Journal of Medical Imaging and Radiation Oncology*, 68(3), 257–264. <https://doi.org/10.1111/1754-9485.13621>
- Temsah, O., Khan, S. A., Chaiah, Y., Senjab, A., Alhasan, K., Jamal, A., & Temsah, M. H. (2023). Overview of early ChatGPT's presence in medical literature: Insights from a hybrid literature review by ChatGPT and human experts. *Cureus*, 15(4). <https://doi.org/10.7759/cureus.37281>
- Topol, E. (2019). *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK.
- Trabilsy, M., Roberts, A., Ahmed, T., Silver, M., Manasseh, D. M. E., Andaz, C., ... Feinberg, J. A. (2023). Lack of racial diversity in surgery and pathology textbooks depicting diseases of the breast. *Journal of Surgical Research*, 291, 677–682. <https://doi.org/10.1016/j.jss.2023.07.019>
- US Food and Drug Administration. (2024). *Transparency for machine learning-enabled medical devices: Guiding principles*. US Food And Drug Administration. <https://www.fda.gov/media/169134/download>
- van der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–214. <https://doi.org/10.3109/0142159X.2012.652239>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.3109/0142159X.2012.652239>
- Weidener, L., & Fischer, M. (2023). Teaching AI ethics in medical education: A scoping review of current literature and practices. *Perspectives on Medical Education*, 12(1), 399–406. <https://doi.org/10.5334/pme.954>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- Zhu, S., Wang, W., & Liu, Y. (2024). *Quite Good, but Not Enough: Nationality Bias in Large Language Models--A Case Study of ChatGPT*. arXiv preprint arXiv:2405.06996. <https://doi.org/10.48550/arXiv.2405.06996>