Physician Readiness for Artificial Intelligence in Clinical Practice

An Analysis of the Medical Education Landscape

September 2026

COUNCIL OF MEDICAL SPECIALTY SOCIETIES

Project Partners & Acknowledgements

As a coalition of more than 50 organizations representing nearly 1 million physicians, the Council of Medical Specialty Societies (CMSS) elevates the unique role of medical specialty societies. We advance the expertise and collective voice of specialty societies and the patients they serve to drive meaningful change in the future of healthcare.

This report seeks to map US physicians' Artificial Intelligence (AI) literacy and challenges in education and workflow, identify factors impacting AI's use in patient care, and support our member societies and the broader community in preparing physicians for AI-enabled practice across the medical education continuum.

Given their deep expertise in integrating AI into medical education and clinical practice, CMSS partnered with Cornelius James, MD, FACP, FNAP, and his team at the University of Michigan Medical School to conduct this comprehensive landscape analysis of Physician Readiness for Artificial Intelligence in Clinical Practice. This analysis maps AI literacy and readiness across the continuum of medical education, including Undergraduate (UME), Graduate (GME), and Continuing (CME) Medical Education.

Many thanks to the Project Teams at the University of Michigan Medical School and CMSS for their contributions to this report. Special thanks to Suzanne Miyamoto, PhD, RN, FAAN, and her team at the American Academy of Nursing for their collaboration and to Pete Clardy, MD, at Google for Health for his wise counsel.

University of Michigan Medical School

Cornelius A. James, MD, FACP, FNAP

Tony H. Chang, MS

Sanzida Chowdhury, BS

LaTeesa N. James, MLIS, MA

Benjamin Y. Li, BS

Andrew Wong, MD, MS, *Consultant*

Council of Medical Specialty Societies

Helen Burstin, MD, MPH, MACP

Ndifreke Ikpe, MHA

Kristen Ndure

Julia Peterson, CAE

This report was created with support from Google.org. Any findings or conclusions expressed in this report do not necessarily reflect the views of Google.org.

In-text superscript citations are hyperlinked to the corresponding entry in Appendix A (References); hover over a superscript to display the full citation or click the superscript to navigate to the entry.

Table of Contents

1. Background and Purpose of the Analysis

Artificial intelligence (AI) is being integrated into the health system faster than the educational structures meant to prepare physicians to use it. AI-based tools are already being explored across administration, clinical care, and research, and patients are using them to manage their own health. A commissioned review of the medical education literature by Boscardin and colleagues mapped 455 articles across five domains—admissions; classroom-based learning and teaching; workplace-based learning and practice; assessment, feedback, and certification; and program evaluation and research—and concluded that AI has begun to transform the status quo while attention to its limitations and risks remains essential.[1]

This mismatch frames the problem. Clinical adoption is often informal, rapid, and outside institutional oversight, whereas formal curricula, validated competencies, faculty capacity, assessment systems, and governance develop more slowly and unevenly. Demand far outstrips delivery: global student surveys report up to 92% wanting AI integrated into core curricula,[2] and a national physician survey found 92% of practicing physicians wanting more AI training while only 11% of those with any training described receiving “a lot.”[3]

The competencies physicians now need extend beyond tool operation to foundational AI literacy; critical appraisal of AI tools for validity, bias, and real-world applicability; ethical, legal, and governance fluency; and the judgment to collaborate with and override AI while preserving humanistic, patient-centered care. A clinical reality benchmark sharpens the stakes: models scoring 84–90% on medical knowledge benchmarks fall to 45–69% on clinical decision-making and 40–50% on patient safety tasks,[4] a knowledge-practice performance gap corroborated by a systematic review of 39 clinical large language model (LLM) benchmarks.[5] In a systematic review of 519 studies evaluating health care applications of LLMs, only 5% used real patient data and 1.2% reported calibration.[4,6] This evidence highlights the need for frontline physicians to be equipped with the knowledge, skills, and attitudes necessary for effective collaboration with AI, and to critically appraise its outputs. Some have proposed further stratifying clinicians along a consumer–translator–developer continuum, distinguishing clinicians by depth of AI engagement so that curricula can be differentiated by expected role.[7] A parallel stratification appeared earlier in national specialty guidance from the Royal College of Physicians and Surgeons of Canada, which grouped physicians as innovators, early adopters, majority adopters, and late adopters and set out distinct educational needs for each, including, for late adopters, knowing which colleagues to refer patients to when they lack the experience to use AI confidently.[8] Additionally, these competencies are not physician-only: as AI redistributes tasks across care teams, the roles, responsibilities, and accountability of physicians and of the other health professionals they work with will shift together, making interprofessional preparation part of physician AI readiness.

Physician AI readiness. For the purposes of this analysis, physician AI readiness is the capacity of physicians to safely, effectively, ethically, and accountably evaluate, use, supervise, and help improve AI-enabled tools in clinical practice, within appropriate team, institutional, regulatory, and learning health system (LHS) contexts. The construct has three levels. *Individual readiness* comprises the physician’s knowledge, skills, judgment, calibration, critical appraisal of model outputs, and ethical reasoning about AI-informed decisions. *Team/interprofessional readiness* involves the shared workflows, role clarity, communication practices, escalation pathways, and accountability structures needed to use AI safely across a clinical team. *Organizational readiness* comprises the institutional policies, governance, oversight, training infrastructure, quality processes, and LHS feedback loops that support (or constrain) safe clinical use. Physician AI readiness is therefore a multi-level, socio-technical construct rather than a checklist of individual knowledge, skills, and attitudes. While this report focuses on medical education to prepare physicians for AI-enabled clinical practice, readiness at all three levels will be achieved only through the engagement of diverse stakeholders across the health system and society.

Professional identity. Because AI-enabled tools now contribute to diagnosis, decision support, documentation, and patient communication, they touch tasks historically constitutive of the physician role. Two questions organize the professional identity implications for clinical practice and medical education: (1) *What remains uniquely physician work in AI-enabled care?* (2) *How should physicians understand accountability when AI contributes to clinical recommendations or patient care tasks?* These questions treat professional identity as an active educational objective rather than an incidental byproduct, and they recur across the undergraduate medical education (UME), graduate medical education (GME), and continuing medical education (CME) level-specific findings that follow. Interviews with leaders of national organizations conducted for this analysis reinforced a related formation concern: leaders cautioned that learners may use LLMs to outsource their learning rather than as scaffolding for it, displacing the cognitive work through which professional identity is built rather than supporting it.

Cross-sector convergence. The gap this analysis documents is reinforced by convenings that span clinical practice and education. The American Medical Association–Digital Medicine Society (AMA–DiMe) clinical practice landscape report,[4] a 2023 National Academies of Sciences, Engineering, and Medicine workshop on AI in health professions education,[9] the proceedings of the 2024 Josiah Macy Jr. Foundation Conference on AI in Medical Education,[10] the 2025 Millennium Conference ("Artificial Intelligence: Prompts, Hallucinations and the Future of Medical Education"), hosted by the Carl J. Shapiro Institute for Education and Research at Harvard Medical School and Beth Israel Deaconess Medical Center,[11] and a 2026 National Academy of Medicine (NAM) clinicians forum on AI in health and medicine[12] converge on a single conclusion: AI has entered clinical practice (through both clinicians and patients) faster than education and governance structures can prepare for it, while human accountability, judgment, and trust remain the physician's defining responsibilities. Interviews conducted for this analysis affirmed that convergence and extended it to team design: leaders framed the operative questions as which unmet patient needs AI-enabled care teams should be built to address, and how AI-based systems will function as team members alongside clinicians rather than as tools each clinician uses alone.

Connected paradigms — precision medicine, the LHS, and precision medical education. Physician AI readiness is often framed as a standalone problem, but it sits at the intersection of three paradigms that are usually described separately yet share a common logic and a common dependence on a prepared physician workforce. *Precision medicine* (PM) reframes clinical care around data, algorithms, and molecular and computational tools tailored to the individual patient; its foundational blueprints anticipated that realizing it would require physicians trained to interpret and act on these tools, explicitly charging medical education oversight bodies with building the necessary competencies.[13] The *LHS* is the infrastructure and culture that operationalizes this ambition, a continuous cycle in which routine clinical data generate new knowledge that is translated back into improved care, a cycle that AI can now accelerate.[14] This cycle also carries a direct implication for education: the new data and knowledge it produces should inform how physicians are trained, and physicians themselves play a central role in generating, curating, and managing the clinical data on which the LHS depends. *Precision medical education* (PME) also applies the data-to-knowledge-to-practice loop to the physician's own development, using AI to tailor training, assessment, and coaching to the individual learner.[15] Continuing education leaders interviewed for this analysis described the same opportunity from the accreditation side, observing that AI creates a rich data environment in which customized, gap-analyzed learning journeys could be mapped for individual physicians at the point of care.

These are not parallel curiosities but one learning cycle observed at different scales: PM is the clinical aim, the LHS is the system that makes care continuously learnable, and PME turns that same loop inward on the clinician. An LHS cannot safely deploy adaptive algorithms without clinicians who can appraise, calibrate, and remain accountable for the care informed by them, which frames data readiness and then clinician readiness as sequential prerequisites for system-level AI readiness.[14] PM cannot become routine without a workforce fluent in interpreting model-derived clinical information.[13] PME models the workflow-embedded, feedback-driven learning that continuing education for practicing physicians will itself require.[15] Framing physician AI readiness within these connected paradigms brings into focus the target this report returns to throughout: not tool operation, but the judgment to participate in, and be accountable within, a health system that learns.

Environmental impact. One further consideration is only beginning to enter this conversation: the resource cost of the tools physicians are being trained to use. Data centers accounted for 4.4% of US electricity consumption in 2023 and are projected to consume between roughly 7% and 12% by 2028, and their cooling requirements draw heavily on water.[16] Individual clinical uses are small but accumulate; one modeled estimate placed the annual carbon footprint of an AI patient messaging tool across a large academic health system at approximately 48,000 kg of carbon dioxide, equivalent to the amount roughly 2,300 trees remove from the atmosphere in a year, with implementation choices such as model size, prompt length, and data center efficiency altering that figure severalfold.[17] Because clinicians and programs decide which tools to use, how often, and for what tasks in at least some settings—in others, those choices are made institutionally, through platform selection and use restrictions—the environmental consequences of AI sit alongside accuracy, cost, and safety in the judgment physicians are being prepared to exercise. No study identified in this analysis teaches or evaluates this content, so it is treated here as an emerging consideration rather than a scored domain, with implications for training noted in Sections 4.5 and 8.2.

Purpose. This analysis integrates three completed level-specific syntheses (UME, GME, and CME) into a single landscape analysis serving three aims: (Aim 1) *characterize current efforts to prepare physicians for AI-enabled practice across the medical education continuum*; (Aim 2) *elucidate current gaps*; and (Aim 3) *synthesize these into a comprehensive picture for strategic focus and investment*. To those syntheses the analysis adds primary data collected for this project: structured surveys of residency and fellowship program directors and of medical specialty society continuing professional development (CPD) directors, together with interviews of leaders of national organizations spanning the education continuum. These data are used throughout to test, quantify, and extend what the published literature shows; the methods are documented in Appendix I.

Finally, we acknowledge that there are interventions and programs designed to prepare physicians for AI in clinical practice across the continuum of medical education that may not be captured in this landscape analysis for various reasons (e.g., not represented in published literature, under development).

2. Executive Summary

The state of the field. Across UME, GME, and CME, physician AI-readiness education is accelerating but is in its early stages: broad in rationale, frameworks, needs assessments, and small pilots, but thin in implemented, validated, outcome-bearing programming and initiatives. On the 0–4 field-level scale (0 = Absent; 1 = Emerging; 2 = Developing; 3 = Established; 4 = Evidenced), no reporting domain at any level reaches Evidenced, and the landscape clusters at Emerging to Developing.

What is most developed. AI literacy, critical appraisal, and ethics/algorithmic-bias content are the best-represented domains at every level. At the UME level, AI literacy is the most developed domain, driven by widening institutional adoption and repeated short-term knowledge-gain pilots. At the GME level, delivered residency curricula (AI-RADS, an AI curriculum for radiology residents,[18] and the Data-Augmented, Technology-Assisted Medical Decision Making [DATA-MD] curriculum in internal medicine[19]) provide concrete implementation evidence, including measured knowledge gains, though the evidence base remains heavily radiology-weighted. At the CME level, professional society guidance is the most mature accreditation-adjacent sub-domain, reaching Developing (2), although the consolidated accreditation, licensing, board examination, society guidance, and continuing certification (CC) domain remains Emerging (1). Across levels, competency frameworks converge on a common core: foundational AI and data literacy, critical appraisal of AI outputs, safe clinical integration, ethics, equity, and governance (including algorithmic bias), and human–AI collaboration with attention to professional identity.[20]

What frontline educators report. Primary data collected for this analysis converge with the published evidence and reinforce it. Among 114 residency and fellowship program directors surveyed through the Organization of Program Director Associations (OPDA), only 9.6% reported that their residents were not using AI tools; use was most common for clinical documentation and administrative tasks, including ambient scribes (67.5%), and through commercial LLMs (57.9%). Yet 72.8% were only slightly confident or not confident that their residents can appraise AI outputs for accuracy, bias, and patient safety, and only 2.6% were very confident. Among 34 CPD directors representing 28 medical specialty societies, 20.6% offered dedicated AI-focused programming for CME or CC credit, 29.4% had no active AI-focused programming at all, and 64.7% had not yet formally assessed member interest in AI content. Both groups named the same two leading barriers: limited faculty expertise (50.0% of program directors; 64.7% of CPD directors) and the absence of standardized competencies (45.6% and 58.8%, respectively). These findings corroborate the gap structure described below from the vantage point of those responsible for delivering the education.

The cross-cutting gaps. Seven gaps span the medical education continuum: (1) no behavior-change or patient-outcome data, with reported outcomes stopping at reaction and knowledge (Kirkpatrick Levels 1–2); (2) validated AI-competency assessment instruments are essentially absent, with reliance on non-validated self-report; (3) AI-specific milestones/entrustable professional activities (EPAs) exist only as proposals; (4) no binding accreditation standard requires AI, and AI is largely absent from licensing/board examinations; (5) faculty development is a near-universally named but largely unmet need; (6) reporting standards, dissemination, and scale-up infrastructure are underdeveloped, hindering pooled learning and the spread of effective educational interventions;

and (7) institutional and program-level governance of AI use, the policies and oversight that determine how trainees and physicians may use AI in clinical work, is sparse and rarely specific to medicine or clinical practice.

Readiness drops across the continuum. Implementation evidence is most visible in UME (literacy teaching, national curricular tracking) and in pockets of GME (radiology especially), then thins at the practicing physician level, where appetite may be highest but delivery is heterogeneous and largely unevaluated. Evidence shows practice has already moved ahead of education: AI is embedded in workflows and clinical encounters while the educational evidence base remains largely conceptual and pre-implementation.

The Association of American Medical Colleges (AAMC) New and Emerging Areas competency series (currently in draft form and expected to be released in October 2026) is the first organization-led, cross-continuum, tiered physician AI competency framework, narrowing but not yet closing the coordination gap that keeps defined competency frameworks at Developing.[20] The International Advisory Committee for Artificial Intelligence (IACAI) UME and GME matrices add organization-backed guidance;[21] the Macy landscape review,[1] innovator interviews,[22] and conference recommendations[11] supply continuum-wide action agendas; and a UME curriculum[23] and accredited CME offerings document real provision. All remain drafts, guidance, landscape maps, or course inventories without effectiveness data, so they reinforce rather than reset the gap structure.

Top priorities for investment. Four priorities follow from the gaps: (1) build and validate AI-competency assessment: instruments, milestones/EPAs, and outcome measures reaching behavior and patient levels; (2) fund and deliver faculty development rather than only naming it; (3) align accreditation, licensing, certification, and CC with AI competencies; and (4) adopt and operationalize a shared cross-continuum competency framework.

3. Undergraduate Medical Education (UME)

3.1 Narrative Overview

The consolidated UME corpus is broad but shallow: rich in rationale,[24,25,26,27,28] frameworks, needs assessments, and small pilots, and thin in implemented, validated, outcome-bearing programming. Demand is near-universal while delivery is sparse: surveys find strong student appetite, with roughly two-thirds to over 90% of students endorsing formal AI training across national and international surveys[2,29] alongside minimal current teaching and assessment. Organized around the readiness domains, AI literacy is the best-developed domain, the only domain with both breadth and outcome data; clinical application and critical appraisal are specified through structured appraisal scaffolds such as the VALIDATE checklist for trainee appraisal of AI-generated output[30] and the VERIFY framework.[31] A distinct educational innovation, DEFT-AI, adapts an established model of clinical supervision — Diagnosis, Evidence, Feedback, Teaching (DEFT) — so that supervisors can probe, coach, and give feedback on how a learner used an AI tool in a specific patient encounter.[32] Neither these appraisal scaffolds nor the supervision model has yet been evaluated for effectiveness. Ethics, equity, legal, and governance content, including algorithmic bias, is among the most uniformly emphasized and is anchored to society and policy instruments; human–AI collaboration and professional identity/systems thinking recur as cross-cutting themes with consistent warnings against deskilling and overreliance and a shared “augment, not replace” framing;[33,34] and faculty readiness is repeatedly identified as a rate-limiting gap. At the organizational level, the AAMC New and Emerging Areas in Medicine competency series for AI—six cross-continuum domains across three developmental tiers (Novice, Proficient, Expert)—gives UME an organization-backed competency target that spans the continuum, though it remains in draft, with release expected in October 2026.[20]

Readiness needs also differ within UME by phase, and the sections below apply both a pre-clerkship lens (foundational literacy, appraisal habits, and data and privacy basics) and a clerkship lens (supervised application in patient care). Pre-clerkship students encounter AI primarily as a learning tool, while clerkship students increasingly meet it embedded in clinical information seeking: in a national survey of Canadian medical students, pre-clerkship students used LLMs more often for coursework and objective structured clinical examination (OSCE) preparation, while clerkship students used them more often for clinical questions.[29]

On educational approach, current efforts cluster around short, modular electives and asynchronous modules: a four-week elective,[35] a 30-minute module,[36] a mandatory Belgian program,[37] a Romanian course,[38] a 100-hour “hacking-education” course,[39] and trainee-led datathons,[40] with broad agreement that the desired end-state is longitudinal, scaffolded, spiral integration into existing organ-system blocks and clerkships rather than “bolt-on” courses. Assessment is limited: overwhelmingly formative and self-report, with validated competency instruments scarce and AI-specific milestones/EPAs existing only as proposals.

3.2 What the Literature Shows

The strongest sources are large, rigorous reviews (a Best Evidence Medical Education [BEME] Guide of 278 publications,[33] a meta-ethnography spanning ~48 countries,[34] and additional systematic, scoping, and narrative reviews[41,42,43,44,45]) and a small number of consensus studies; these consistently conclude that frameworks, demand, and rationale advance far faster than implementation, assessment, and evaluation. Delivered literacy courses report significant short-term knowledge or confidence gains (Krive 61%→97%;[35] Agarwal 74%→87%, P<.001;[36] Rezazadeh all measured outcomes P<.001[39]), and national AAMC data document rapidly widening adoption (schools reporting AI in the curriculum rising from 88 of 166 responding US schools in 2022–2023 to 140 of 182 responding US and Canadian schools in 2023–2024, or 53% to 77%, a 59% increase in a single year; AI also moving into required curricula, from 61 of 88 schools to 107 of 140; basic AI concepts the most common required topic at 73%) while governance lags (47% with an appropriate-use policy; 30% providing learners secure access to an AI model[46]). Evidence quality is otherwise dominated by single-institution, self-selected, self-report studies; no domain reaches Evidenced. The most recent learner-side data highlight this mismatch: in a national survey of Canadian medical students fielded in late 2025, 96.5% reported using at least one LLM and 90.2% had encountered inaccurate output, yet only 11% had received formal instruction on LLM use and 67.7% agreed that schools should integrate formal training.[29]

The Stanford "AI in Clinical Practice" curriculum is a concrete, competency-mapped UME outline spanning foundational knowledge, integration skills, ethics/legal, and critical appraisal of model performance.[23] This provides a named exemplar consistent with the corpus pattern of single-institution outlines without reported outcomes. Additionally, as of this writing, a multi-institutional group led by the Carl J. Shapiro Institute for Education and Research at Harvard Medical School and Beth Israel Deaconess Medical Center has begun work on a five-module UME curriculum for pre-clerkship students addressing AI concepts, ethics, clinical skills, professionalism, and AI limitations.[11] Further, the IACAI Matrices (v2026)[21] provide structured, international organization-backed frameworks across AI-integration domains and role levels, reinforcing framework breadth while remaining guidance rather than an adopted, validated, outcome-bearing standard.

Specialty-specific signal (generalizability caution). Specialty content is heavily imaging-concentrated, with psychiatry, neurosurgery, and primary care examples typically represented by single studies; reviews note a large share of work specifies no discipline, flagging a generalist gap, so imaging-derived findings should not be extrapolated across UME.[47,48,49,50]

3.3 UME Gap-Analysis Matrix

Consolidated six-domain matrix — UME

Reporting domain	Score (codes)	Basis for the consolidated score
1. Defined competency frameworks	**2 (L, P, G)**	Multiple frameworks converge on a common core of competencies; none are standardized or accreditation-adopted.[47,51,52,53] **Developing**.
2. Formal curricula	**2 (L, G)**	Literacy is the sub-domain with knowledge-gain pilots;[35,36,54] application, appraisal, and specialty content are specified but inconsistently delivered and rarely assessed.[55,56] **Developing.**
3. Faculty development	**1 (L, G)**	Qualified faculty and sustained resourcing are documented barriers, not delivered programs;[57,58,59] interdisciplinary collaboration is demonstrated but project-specific.[60] **Emerging**.
4. Validated assessment methods	**1 (L, P)**	Validated instruments and AI-specific milestones/EPAs are aspirational;[55,61] outcomes stop at Kirkpatrick Levels 1–2 with no behavior or patient data.[51,59] **Emerging**.
5. Ethics / equity / legal	**2 (L, P, G)**	Ethics, governance, and algorithmic bias are among the best-represented content areas and society-anchored;[52,62] institutional policy adoption is growing but generic and lagging.[46,62] **Developing**.

Reporting domain	Score (codes)	Basis for the consolidated score
6. Accreditation / licensing / board exam / society guidance / CC	**1 (L, P, G)**	No binding accreditation standard requires AI and AI is largely absent from licensing/board exams;[46,63] society guidance is abundant but non-binding and fragmented; no CC/AI pathway operates.[47,64] **Emerging.**

Source codes: L peer-reviewed literature · P policy/accreditation · G gray literature/society guidance · O other.

The detailed 18-sub-domain matrix for UME appears in Appendix C.

3.4 Gaps at the UME Level

UME shows the seven continuum-wide gaps synthesized in Section 7: absent behavior- and patient-level outcomes, scarce validated assessment, proposal-only milestones/EPAs, no binding accreditation or licensing requirements, named-but-unmet faculty development, and underdeveloped reporting and dissemination infrastructure, compounded by barriers to sustained resourcing. As of the writing of this report, the Liaison Committee on Medical Education (LCME) is expected to release an academic year 2027–2028 update to Standard 7.2 that will require clinical experiences, education, and experiential learning that include understanding the appropriate use of AI and other emerging technologies in diagnosis and patient management. The evidence base is geographically skewed toward North America and Europe, with recurrent equity concerns for low- and middle-income settings. A further, largely unexplored gap concerns sequencing and prerequisite knowledge: the literature has not yet defined what foundational knowledge students should acquire before using AI, and particularly LLMs, to augment clinical reasoning; when in training LLMs or ambient scribes should first be introduced into patient care; or which traditional knowledge and skills matter less, and which matter more, in AI-augmented practice. These remain open questions in the medical education community. The seventh of these gaps, institutional governance of AI use, is especially thin at the UME level: audits of US medical school AI documents find that most institutions lack a formal, medical-student-specific AI policy, with available materials emphasizing academic integrity over the oversight and accountability structures that would govern clinical AI use.[62,65]

4. Graduate Medical Education (GME)

4.1 Narrative Overview

The GME evidence base is early-stage and uneven. GME-relevant work is dominated by expert commentaries, competency-framework proposals, scoping and systematic reviews of heterogeneous literature, and single-institution needs assessments or perception surveys,[66,67] with only a small number of implemented, evaluated residency curricula. The few GME curricula with primary outcome data—AI-RADS for radiology residents,[18] the DATA-MD curriculum for internal medicine residents,[19] the radiology residency curricula covered by a systematic review,[68] and a multidisciplinary AI ethics workshop[60]—are uniformly single-site, small-N, and report only Kirkpatrick Level 1–2 outcomes. Separately, Yale's Artificial Intelligence in Clinical Care module is an established, web-based curriculum developed within Yale's primary care pediatrics program and adaptable by residency programs in pediatrics, internal medicine, and combined internal medicine–pediatrics. It does not itself report primary outcome data, though the developers are in the process of gathering it, and is noted here as an available resource rather than as outcome evidence.[69] Mixed-level mapping places GME in the minority of the evidence base,[33,59] specialty representation is dominated by radiology, and geography skews to North America and Europe.

Across the readiness domains, AI literacy is the most consistently addressed and critical appraisal carries the best effectiveness signal; clinical application is mostly proposed or described rather than systematically delivered; ethics/governance and algorithmic bias are near-universally identified but largely conceptual and rarely assessed; human–AI collaboration recurs through automation bias and overreliance concerns; and faculty capacity is consistently flagged, with an "inversion of expertise" in which trainees may outpace supervising faculty.[32,67] The appraisal scaffolds and supervision model introduced at the UME level are also relevant here: the VALIDATE checklist and the VERIFY framework offer stepwise approaches to appraising AI-generated output,[30,31] and DEFT-AI, which adapts the Diagnosis, Evidence, Feedback, Teaching model for clinical supervision of AI use, is the most directly workplace-based of the three and therefore particularly applicable to residency training.[32] As at the UME level, no published outcome data are yet available for either the appraisal scaffolds or the supervision model.

Implemented approaches favor short didactic-plus-hands-on formats; for example, lectures with journal clubs, short module series, and workshops, with calls for longitudinal integration, protected time, and interdisciplinary teaching teams. Assessment is predominantly formative and self-reported, and AI-specific Accreditation Council for Graduate Medical Education (ACGME) milestones or EPAs are not in place. Here too, the AAMC New and Emerging Areas in Medicine draft competency series for AI supplies an organization-backed (though still draft) competency target, six cross-continuum domains across three developmental tiers, against which GME programs can align emerging milestones and EPAs.[20]

4.2 What the Literature Shows

The strongest GME evidence is narrow: one curriculum (DATA-MD) supplies primary quantitative outcome data with effect sizes—that is, the only study to report comparative effect sizes rather than Kirkpatrick Level 1–2 reaction or knowledge outcomes alone;[19] one systematic synthesis covers radiology resident curricula;[68] and a scoping review covers pediatric GME.[70] Additionally, in a scoping review of AI-related educational programs, 17% (5/30) were at the GME level, and all of these were radiology-focused.[55] Direct resident survey evidence reinforces the training gaps: among internal medicine residents across five North Carolina programs, 26% reported professional LLM use, none had received formal training, 94% expressed interest in learning more, and 87% believed residencies should, or may need to, offer a formal curriculum, with differential diagnosis, research, and self-directed learning the leading uses.[71] Comparable national data from Canada predate the generative-AI era and indicate that the training gap is long-standing: in Royal College of Physicians and Surgeons of Canada task force surveys, 42% of Resident Affiliates and 41% of Fellows rated themselves as having no or very little familiarity with AI, nearly 70% of residents reported no or very limited exposure to AI in their work or education environment, and 85% believed they had not received adequate teaching to prepare them to use AI in future practice, even as 59% of residents expressed high interest in AI-related education.[8]

The IACAI GME stakeholder mapping (Tables 1–2; a domains-by-stakeholders needs/roles matrix and a RACI [responsible, accountable, consulted, informed] engagement matrix)[21] structures responsibilities across micro/meso/macro/mega stakeholders and is explicitly illustrative rather than exhaustive; it enriches the governance and stakeholder-engagement narrative without adding delivered curriculum or outcome evidence. Boscardin and Gin's national landscape review and innovator interviews[1,22] add continuum-wide evidence directly relevant to workplace-based GME: AI scribes, clinical decision support, and machine vision interpretation as documented use cases, alongside concerns about deskilling, monitoring fatigue, and validation, reinforcing existing Developing-level scores and the absence of higher-order outcomes.

Recency and skew (generalizability caution). GME implementation evidence clusters in radiology; other specialties are typically single studies. Some perception data predate widespread generative AI, and some curricula were developed pre-GPT and did not include generative AI, underscoring the need for continual updating as AI technologies continue to advance. This currency limitation is not unique to GME; it recurs across the UME, GME, and CME literature, where rapidly evolving AI technologies can outpace both curricula and their published evaluations.

4.3 Sentiments / Experiences of Experts and Stakeholders

Residents and program leaders express strong demand for structured AI training and supervised clinical exposure, tempered by concerns about adequate supervision, valid assessment, accountability for AI-assisted work, and uneven access across specialties. Consistent with the published literature, program-level readiness—faculty capacity, protected time, and resourcing—is repeatedly identified as the principal constraint.[22,66,67]

Program-director survey evidence (primary data). A structured survey of residency and fellowship program directors conducted for this analysis through the Organization of Program Director Associations (OPDA; N = 114) documents how far routine clinical use has outpaced formal curriculum. Only 9.6% (11 of 114) reported that their residents were not currently using AI tools. Use was most commonly reported for clinical documentation and administrative tasks, including ambient scribes (67.5%, 77 of 114), followed by commercial LLMs used for clinical decision support (57.9%, 66 of 114) and models embedded in the electronic health record (32.5%, 37 of 114) (Figure 1).

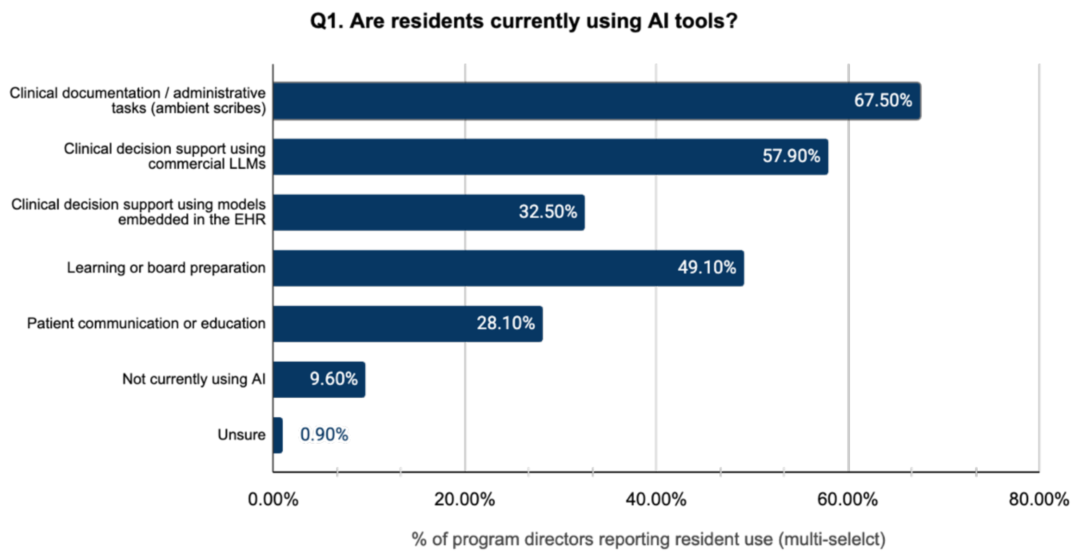

***Figure 1**. Resident AI use by category (Q1, n = 114; select all that apply). Total exceeds 100%.*

The barriers program directors identified are structural rather than attitudinal: limited faculty expertise (50.0%, 57 of 114), lack of standardized competencies (45.6%, 52 of 114), and limited evidence about effective educational approaches (41.2%, 47 of 114) (Figure 2).

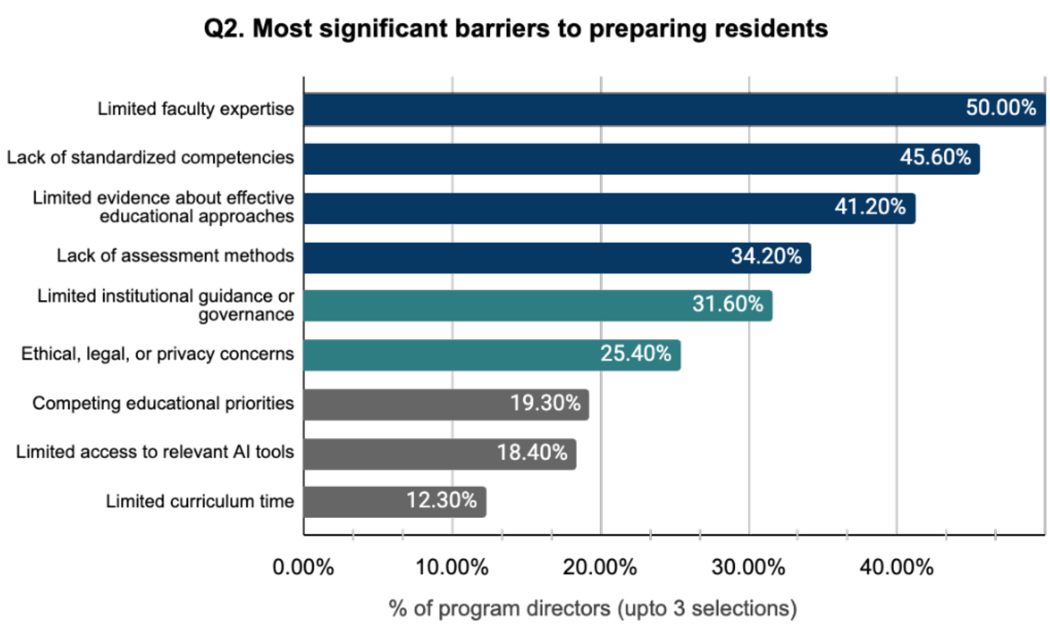

***Figure 2**. Most significant barriers to preparing residents for AI use (Q2, n = 114; up to three selections).*

Asked what training matters most for AI-enabled practice, program directors selected recognizing the limitations of AI, its potential for error, and bias (85.1%, 97 of 114) and critical appraisal of AI outputs (68.4%, 78 of 114) above all other options (Figure 3). When asked what training would most improve safe and critical use, respondents converged on critical appraisal (19%) and critical thinking (15%) as content, and on case-based teaching and asynchronous modules as format; direct observation was named by only two respondents. Asked what must be in place first, respondents named protection against deskilling—that is, safeguarding core clinical skills (10.6%)—and assessment methods (10.6%).

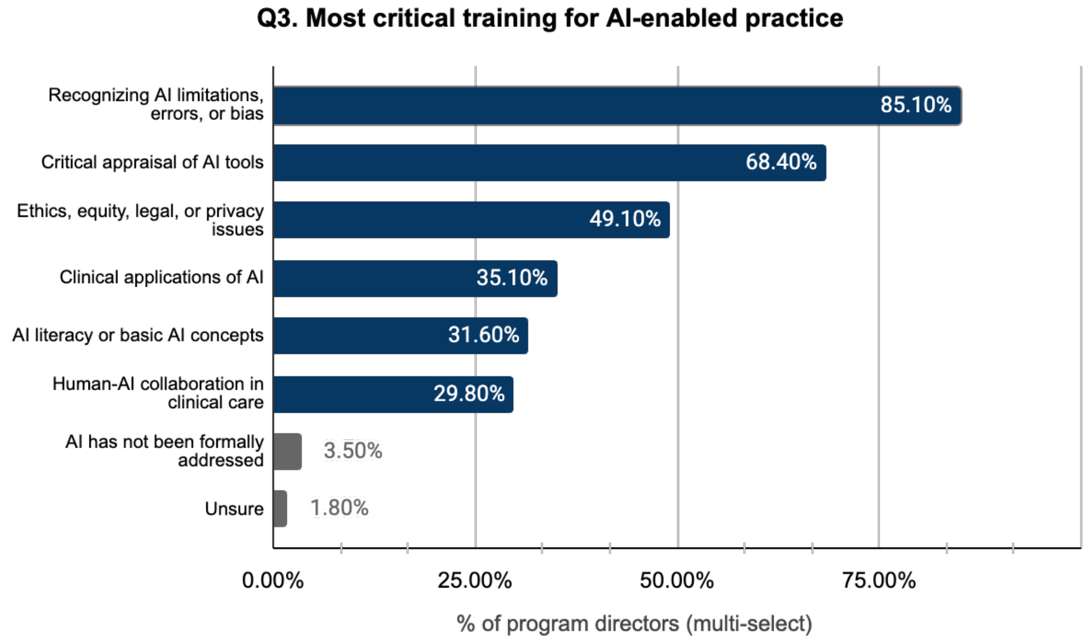

Figure 3*. Most critical training for AI-enabled practice (Q3, n = 114; select all that apply).*

Confidence in residents' appraisal skills is correspondingly low: 72.8% (83 of 114) were only slightly confident (38.6%, n = 44) or not confident (34.2%, n = 39) that their residents can appraise AI outputs for accuracy, bias, and patient safety; only 2.6% (3 of 114) were very confident (Figure 4).

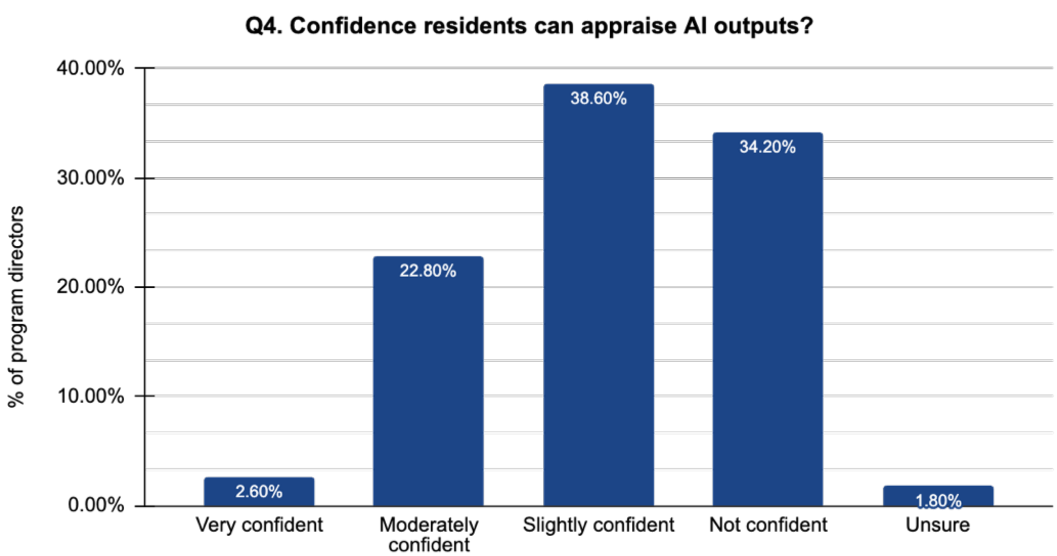

Figure 4. *Program directors' confidence in residents' ability to appraise AI outputs (Q4, n = 114).*

Free-text responses centered on accountability. Program directors stated repeatedly that residents must retain ownership of the clinical note, that editing AI-generated text and checking it for hallucination is obligatory rather than discretionary, and that graduating residents should be willing to use tools such as ambient scribes to reduce administrative burden while maintaining a critical eye on whether outputs are accurate, evidence-based, and ethical. Faculty development was described as a prerequisite rather than an adjunct: programs cannot teach or assess safe AI use if the educators themselves do not understand how to implement it responsibly. On delivery, respondents were explicit that didactics alone are insufficient, asking instead for interactive, case-based, and workshop-based sessions that demonstrate safe use in real time in the clinical environment.

4.4 GME Gap-Analysis Matrix

Consolidated six-domain matrix — GME

Reporting domain	Score (codes)	Basis for the consolidated score
1. Defined competency frameworks	**2 (L, G)**	GME-relevant frameworks (e.g., an ACGME-milestone-style radiology framework; an AIFM-ed family-medicine framework mapped to CanMEDS-FM) remain proposals or single-setting; 45.6% (52 of 114) of program directors surveyed for this analysis cite the absence of standardized competencies as a leading barrier.[52,72,73] **Developing**.
2. Formal curricula	**2 (L, G, P)**	Literacy, application, appraisal, and specialty content appear in implemented residency curricula (AI-RADS; DATA-MD) but are single-site and mostly radiology-concentrated;[18,19,68] an ambient AI-scribe pilot is the most concrete deployment.[74] **Developing**.
3. Faculty development	**1 (L, G)**	An inversion of expertise (trainees outpacing supervisors) and low attending role-modeling are documented;[32,34] in the OPDA program-director survey conducted for this analysis, 50.0% (57 of 114) named limited faculty expertise as a leading barrier to preparing residents for AI-enabled practice. No evaluated faculty development program exists, and interdisciplinary teaming appears but is not formalized.[22] **Emerging**.
4. Validated assessment methods	**1 (L, P)**	Validated instruments are largely absent and milestones/EPAs are proposed, not in force; outcomes are Level 1–2 and single-site (DATA-MD effect sizes 0.74–1.80; AI-RADS gains); program directors surveyed for this analysis named assessment methods among the guardrails that must be in place before AI training can be delivered safely.[18,19,55,72] **Emerging**.
5. Ethics / equity / legal	**2 (L, G)**	Governance, regulation, and bias are recurrently taught (DATA-MD modules; European Society of Gastrointestinal Endoscopy [ESGE] guidance; psychiatry regulatory mapping) but weakly assessed: DATA-MD's ethics/law module showed no measurable gain.[19,75,76] Evidence on the governance of residents' own clinical AI use is limited: a review of five major medical and specialty organizations found that none had issued resident-specific guidance.[77] **Developing**.
6. Accreditation / licensing / board exam / society guidance / CC	**1 (P, L, G)**	No binding accreditation standard requires AI; the clearest concrete signal is the 2021 addition of AI by the American Board of Radiology (ABR) to its Qualifying (Core) Exam noninterpretive-skills section;[68] society guidance is advisory and uneven; no CC pathway.[76,78] **Emerging**.

Each GME score reflects the GME-specific weight of evidence; where mixed-level evidence ran higher, the score is held to the GME-specific strength. Primary program-director survey data collected for this analysis (source code O; Appendix I) corroborate the bases above but do not change any consolidated score.

The detailed 18-sub-domain matrix for GME appears in Appendix C.

4.5 Gaps at the GME Level

GME shows the same seven gaps, with implementation and evaluation lagging well behind framework development in every domain. The evidence remains heavily radiology-weighted, single-site, and concentrated in North America and Europe; licensing/board AI content is essentially limited to a single concrete example, the 2021 addition of AI by the ABR to the noninterpretive-skills section of its Qualifying (Core) Exam.[68] The governance gap is equally visible here: a review of five major medical and specialty organizations found no resident-specific guidance on AI-assisted clinical documentation, leaving only general, non-binding principles.[77] The sequencing and prerequisite knowledge gap noted at the UME level applies equally in GME: the literature has not defined what foundational competencies residents should hold before supervised clinical use of AI, or how AI training should be sequenced across postgraduate training.

Interviews conducted for this analysis add an accreditation-timing dimension to this governance gap. Leaders in GME accreditation described the difficulty of writing Common Program Requirements (CPRs) for AI, given the concern that any requirement specific enough to be useful risks being outdated soon after it is written, and indicated that new standards are anticipated to be introduced in September 2026 and to take effect in 2028. Whatever those standards ultimately contain, the interval itself is consequential: it leaves sponsoring institutions and individual programs responsible for their own interim governance and usage policies during precisely the period in which resident AI use is becoming routine, which raises the priority of local policy development over deferral to a forthcoming national standard.

A further consideration, absent from every competency framework and curriculum identified at this level, is the environmental cost of the tools that residents use. Trainees interact with AI predominantly at the inference stage, where the resource demand of a single task is low but the frequency of use makes the aggregate substantial. A 2026 graduate medical education editorial argues that environmental stewardship should be among the metrics by which institutions and programs judge responsible AI use, points to practical levers such as institution-wide platforms that provide transparency about energy use and emissions, and sets out questions the community has yet to answer: whether governance structures for AI account for environmental impact, what tools programs would need to measure it, and where the topic belongs within teaching on medical ethics.[16] Nothing in the reviewed GME evidence addresses this content, so it is recorded here as an open area rather than a scored gap.

5. Continuing Medical Education (CME)

5.1 Narrative Overview

The CME evidence base draws on 19 retained CME-relevant sources plus two AMA context documents used for stakeholder sentiment and a clinical reality benchmark. It is predominantly conceptual and pre-implementation: of the 19 sources, only one reports a delivered CME intervention with measured learner outcomes, and the remainder are reviews, commentaries, position statements, framework proposals, or attitudinal surveys. Organized around the readiness domains, AI literacy is the most consistently addressed and the one with the clearest CME signal; clinical application is widely endorsed but rarely delivered and evaluated; critical appraisal is a strong conceptual theme via an evidence-based medicine (EBM) analogy; ethics is among the best developed and the one area where ethics education has been delivered and measured; human–AI collaboration is addressed through cognitive bias and appropriate reliance; professional identity appears largely as commentary; and faculty readiness is the most consistently named barrier and the least addressed. Because the CME audience is largely the same cohort of practicing physicians who serve as clinician-educators, strengthening AI-focused CME may also be a lever for closing the faculty readiness gaps flagged at the UME and GME levels; faculty development and continuing education are, in effect, related problems addressed at different points along the career continuum.

Delivery models are dominated by proposals instead of evaluated programs: short workshops and prompt-a-thons, case-based learning with deliberately introduced AI errors, flipped classroom modules, staged longitudinal roadmaps, and phased society curricula, with interprofessional delivery a recurring recommendation and, in one case, an executed design. Across 30 reviewed programs, none referenced a guiding learning theory, and program evaluations reached only Kirkpatrick Levels 1–2;[55] the closest thing to a shared assessment metric is the Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS), whose validity for measuring change is contested.[48,55]

5.2 What the Literature Shows

On delivery, one interprofessional workshop demonstrates measured literacy and ethics gains (knowledge $4\pm1.4\rightarrow9\pm1.3/10$; privacy/data-security confidence $1.6\rightarrow4.5/5$; both $P<.0001$),[60] and AI fundamentals were taught in roughly 79% of the 30 AI education programs for medical students, residents, and practicing physicians catalogued in a scoping review.[55] Regarding effectiveness, the evidence is shallow and consistent in its ceiling:

no reviewed program reached Kirkpatrick Level 3 (behavior) or 4 (patient outcomes),[55] and a call to move beyond Level 1 itself signals the absence of higher-level data.

Several resources offer American Medical Association Physician's Recognition Award (AMA PRA) Category 1 Credit: the AMA Ed Hub "AI in Health Care Series" (seven modules, up to 3.5 credits), Harvard's "AI in Clinical Medicine" (up to 27 credits), Mayo Clinic's "Practical Applications of AI" (~6.75 credits), the Cleveland Clinic AI Summit, and journal-based CME (NEJM AI, JAMA+ AI); the Accreditation Council for Continuing Medical Education (ACCME) CME Passport provides search infrastructure through which accredited activities can be located. Additional accredited or society-sponsored offerings include the Society of Teachers of Family Medicine (STFM) Artificial Intelligence and Machine Learning for Primary Care (AiM-PC) curriculum, developed for medical students, residents, and practicing primary care clinicians with American Academy of Family Physicians (AAFP) CME credit,[79] the University of Michigan's Data-Augmented, Technology-Assisted Medical Decision Making (DATA-MD) curriculum, released for CME credit in 2024 (up to 3.5 *AMA PRA Category 1 Credits™*), and the American Board of Artificial Intelligence in Medicine (ABAIM) portfolio of CME-accredited introductory and advanced courses in medical AI.[80] Beyond accredited offerings, freely available web-based generative-AI modules, including those from Google and the American College of Physicians (ACP), further broaden accessible, self-directed provision for practicing physicians, but no outcome data are available. The AAMC New and Emerging Areas draft competencies are also relevant at this level: the series articulates its six domains across three developmental tiers (Novice, Proficient, Expert), giving CME an organization-backed competency target it has otherwise lacked.[20] A small number of institutions also offer more advanced training. These include executive courses such as Harvard Medical School's "AI in Health Care: From Strategies to Implementation," and master's-level programs in AI in medicine at institutions including Yale, the University of Louisville, Saint Louis University, the University of Alabama, Thomas Jefferson University, and the University of Florida. Such programs, however, reach only a small fraction of physicians and cannot, on their own, prepare the entire workforce.

Clinical-reality benchmark (Aim 2 context). The AMA–DiMe landscape report supplies the benchmark the analysis requires and widens the gaps: performance falls from 84–90% on knowledge benchmarks to 45–69% on clinical decision-making and 40–50% on patient safety; roughly 60% of pilots fail; only 15–25% of departments move beyond evaluation; and in a 519-study review of LLM health care applications, only 5% used real patient data and 1.2% reported calibration.[4,6] These make concrete the cost of CME's conceptual-only treatment of appraisal and bias. For CME specifically, this benchmark reshapes both the content and the timing of the educational task. On content, because deployed tools perform worse at the bedside than on knowledge benchmarks, the highest-yield CME is not tool operation but physician calibration: critical appraisal, bias recognition, and knowing when to override, precisely the domains the CME corpus treats only conceptually. On timing, physicians are already accountable for AI-informed decisions, so CME may be uniquely positioned to reach clinicians contemporaneously with deployment; this argues for point-of-care, just-in-time formats tied to the tools in a physician's actual workflow, and for CME accreditors and specialty societies to treat AI appraisal as a recurring rather than one-time educational need. Demand-side conditions are already in place, with 92% of physicians wanting more training,[3] and early delivery vehicles exist, from the ESGE curriculum[76] and the accredited offerings catalogued above to multi-stakeholder clinician education efforts such as those of the Coalition for Health AI,[81] although none yet carries behavior- or patient-level outcome evidence.

5.3 Sentiments / Experiences of Experts and Stakeholders

On attitudes, practitioner surveys report strong demand (89.6% of practicing radiologists/residents and 90–94% of hematology/oncology physicians endorsing AI education[82,83]) against an enthusiasm versus readiness gap, with most exposure to commercial or industry-developed LLMs rather than validated clinical tools.

CPD-director survey evidence (primary data). Primary data collected for this analysis quantify that thin footprint at the society level. A structured survey of continuing professional development (CPD) directors, convened as a Professional Peer Group by the Council of Medical Specialty Societies (CMSS) and completed by 34 respondents representing 28 member societies, found that 47.1% (16 of 34) incorporate some AI content into existing programs, but only 20.6% (7 of 34) offer dedicated AI-focused programming for continuing medical education and/or continuing certification (CC) credit, and 29.4% (10 of 34) have no active AI-focused programming at all (Figure 5; Appendix I).

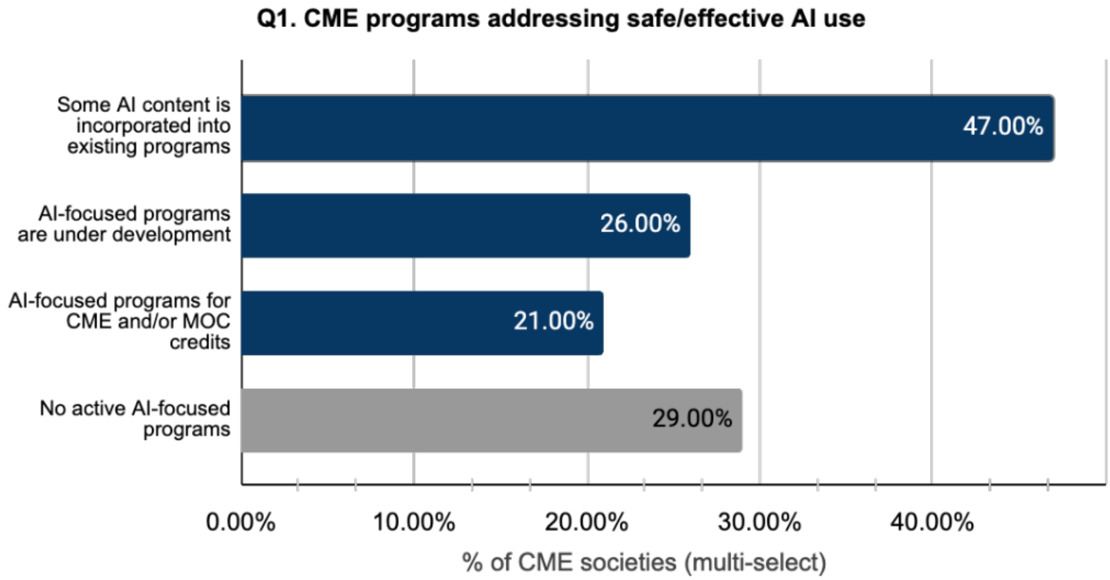

Figure 5. *CME programs addressing safe and effective AI use (Q1, n = 34; single-select).*

Confidence is correspondingly modest: 47.1% (16 of 34) were moderately confident in their society's ability to educate physicians on the safe use of AI, 44.1% (15 of 34) were only slightly confident or not confident, and 8.8% (3 of 34) were very confident (Figure 6).

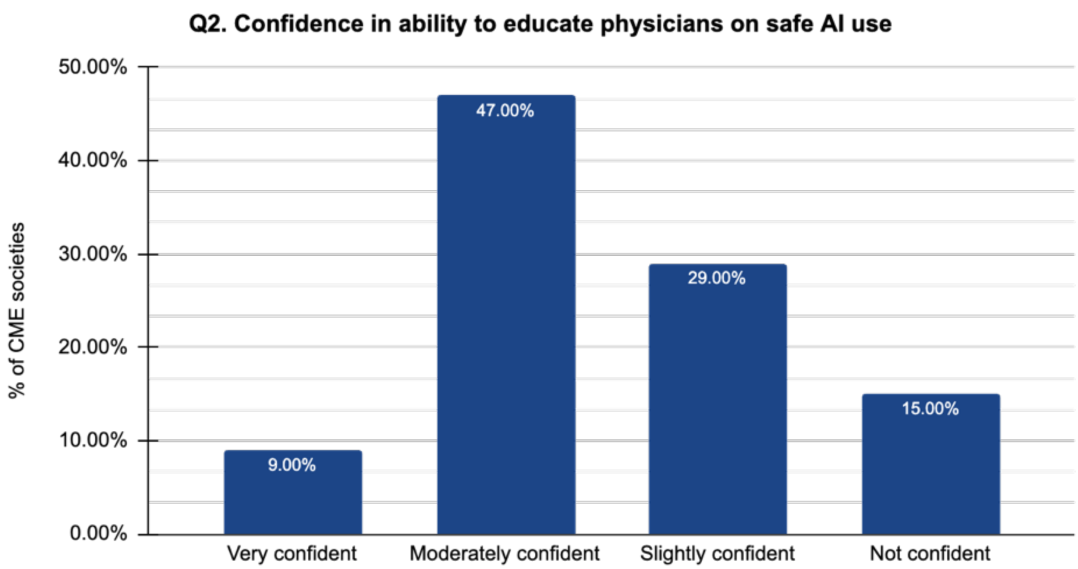

Figure 6. *Confidence in ability to educate physicians on safe AI use (Q2, n = 34).*

Demand-side information is thinner still: 64.7% (22 of 34) had not yet formally assessed their members' interest in AI CME content, so provision decisions are largely being made without a needs assessment.

The barriers societies named echo those reported at the GME level—limited faculty expertise (64.7%, 22 of 34), lack of standardized competencies (58.8%, 20 of 34), and competing educational priorities (52.9%, 18 of 34)—indicating that faculty and competency constraints are continuous across the continuum rather than specific to any one level (Figure 7).

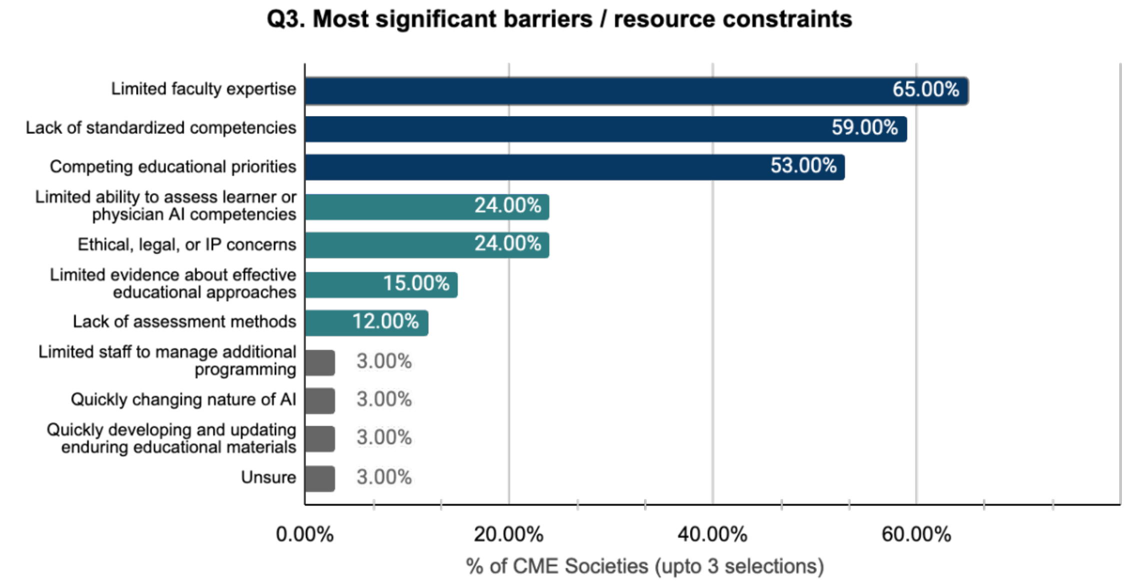

Figure 7. *Most significant barriers / resource constraints (Q3, n = 34; up to three selections).*

Society-level content priorities from the CMSS CPD survey conducted for this analysis align closely with the domains the CME corpus treats only conceptually. Asked what content their members most need, CPD directors ranked ethics, equity, legal, or privacy issues highest (82.4%, 28 of 34), followed by clinical applications of AI (76.5%, 26 of 34) and recognizing AI limitations, errors, or bias (73.5%, 25 of 34). The ordering is notable: the two highest-ranked needs are precisely the appraisal-and-governance domains that the published CME literature addresses in commentary but rarely delivers and almost never evaluates (Figure 8).

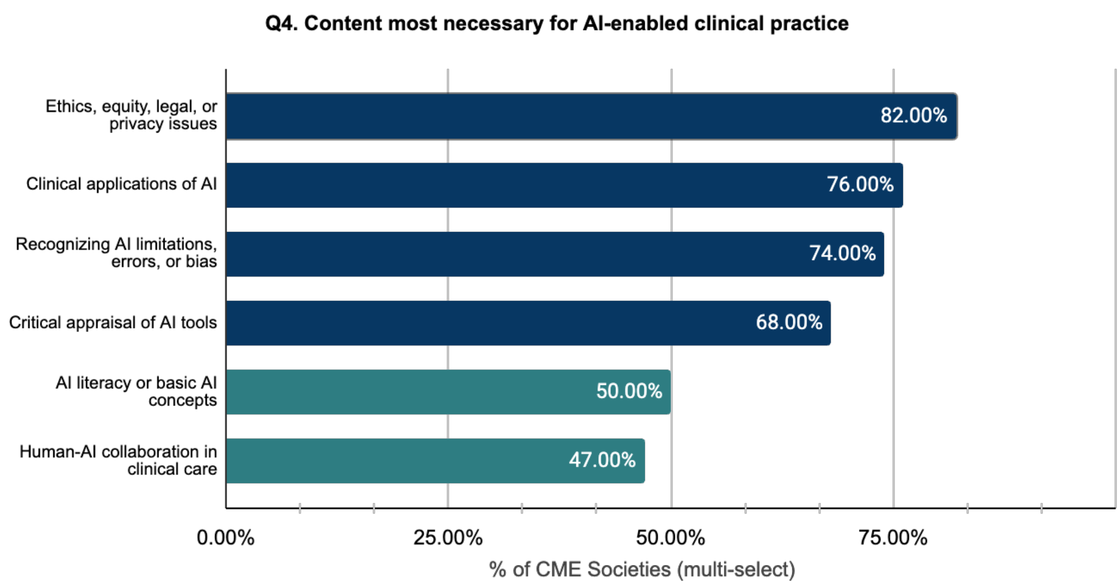

Figure 8. *Content most necessary for AI-enabled clinical practice (Q4, n = 34; multi-select).*

Continuing education leaders interviewed for this analysis paired that opportunity with a caution, emphasizing the need for guardrails around the use of AI in developing CME content itself, even as AI makes it possible to map gap-analyzed, individualized learning journeys rather than leaving physicians to search for CME content on their own.

5.4 CME Gap-Analysis Matrix

Consolidated six-domain matrix — CME

Reporting domain	Score (codes)	Basis for the consolidated score
1. Defined competency frameworks	**2 (L, G)**	Several CME-relevant frameworks are proposed (ESGE statements; primary care domains; Russell's domains; AI-PACE) but diverge and are uncoordinated; only two curriculum frameworks were identified in the reviewed literature; 58.8% (20 of 34) of society CPD directors surveyed for this analysis identify the absence of standardized competencies as a major limiting factor.[47,52,55,76] **Developing.**
2. Formal curricula	**2 (L, G)**	AI literacy/application is the most covered curricular area, with named accredited offerings, but inconsistent in depth and only one delivered study with measured outcomes.[55,60,84,85] **Developing.**
3. Faculty development	**1 (L)**	Faculty readiness is the most consistently named barrier and the least solved (e.g., only 14% of imaging lecturers had formal AI training); collaboration is recommended and once executed but unmeasured; 64.7% (22 of 34) of society CPD directors surveyed for this analysis name limited faculty expertise as a primary barrier to developing AI programming.[47,48,60] **Emerging.**
4. Validated assessment methods	**1 (L, P)**	No widely adopted validated CME AI-competency instrument; milestones/EPAs are structurally more UME/GME-native; only one workshop reports measured learner outcomes; no behavior/patient data.[53,55,60] **Emerging.**
5. Ethics / equity / legal	**2 (L, P, G)**	Policy, governance, and bias are well represented and one ethics workshop showed measured gains, but external policy is cited secondhand rather than embedded as standardized CME.[60,75,84] Governance and policy oversight of practicing physicians' own clinical AI use remains largely unexamined in the literature. **Developing.**
6. Accreditation / licensing / board exam / society guidance / CC	**1 (L, P, G)**	Accreditation/licensing AI mandates and a true CC AI pathway remain absent.[48,55,76,84] **Emerging.**

Scores are anchored in the peer-reviewed CME corpus; the AMA survey (sentiment) and AMA–DiMe landscape report (clinical-reality context, source G) inform context only and do not change any score. Primary CPD-director survey data collected for this analysis (source code O; Appendix I) corroborate the bases above but likewise do not change any score.

The detailed 18-sub-domain matrix for CME appears in Appendix C.

5.5 Gaps at the CME Level

CME shares the continuum-wide gaps: outcome evidence is essentially absent above Kirkpatrick Level 2, validated CME-specific assessment instruments are largely undeveloped, faculty development is a named but unmet need, accreditation and licensing do not yet require physician AI competency, no CC AI pathway is tied to board recertification, and reporting and dissemination infrastructure for effective CME remains underdeveloped.

The defining CME gap is timing. Clinical practice has already moved ahead of CME. AI is embedded in workflows and patient encounters, and the physician's accountable role has shifted, while the CME evidence base remains conceptual and pre-implementation. Provision of AI education and training exists and is growing, yet whether any CME approach changes physician behavior or patient outcomes remains unestablished.

Primary data collected for this analysis identify structural reasons the timing gap persists. Society CPD directors described a mismatch between the pace of AI development and the CME production lifecycle: by the time an activity has moved through planning, development, and review—work that often depends on volunteer subject-matter experts—the tools it describes have changed, so material risks obsolescence before release. Three further constraints compound this. First, even well-designed national education meets local institutional firewalls, because physicians are frequently restricted by their hospitals or health systems as to which AI platforms they may use, creating a disconnect between what societies teach and what learners are permitted to do. Second, societies described an "unknown unknowns" problem: many clinicians do not yet know what they need to know about AI, which suppresses the specific demand signals on which CME needs assessment ordinarily depends. Third, societies reported a tension in their own role, since learners need foundational AI literacy while societies judge their limited resources best spent on specialty-specific clinical considerations, leaving the foundational layer without a clear owner. Some respondents further suggested that clinical application may remain too high-risk for a broad CME push, and that administrative efficiencies, for example within the electronic health record or office-staff workflows, may be a safer starting point for building physician confidence. Respondents also asked for clear policies and best practices governing the use of AI in the creation of educational content itself, a governance question that accreditation and society guidance have not yet addressed.

6. Professional-Organization and Society Guidance

Society and organizational guidance is a relatively mature accreditation-adjacent area in the evidence, but it is advisory, fragmented, and specialty-siloed rather than harmonized into binding standards. The evidence covers, among physician and education bodies, the AMA (augmented-intelligence policy),[86] the AAMC (Principles for the Responsible Use of AI),[87] the NAM (AI Code of Conduct),[88] the World Health Organization (WHO),[89] the Royal College of Physicians and Surgeons of Canada,[8] the Federation of State Medical Boards (FSMB),[90,91] and multiple imaging societies,[48,64,92] with a formal society CME curriculum from the ESGE at the practicing-physician level.[76]

Licensing and national specialty bodies. Two of these instruments bear directly on the accountability and competency questions this report returns to. FSMB policy guidance, adopted by its House of Delegates in 2024, holds that state medical boards regulate licensees rather than tools; that the physician who chooses to use AI is ultimately responsible for that use and accountable for resulting harms; that regulatory scrutiny should rise as an AI tool's function more closely approaches the practice of medicine; and that a physician who cannot fully explain a "black box" output should still be able to offer a reasonable interpretation of it and justify why following or departing from it meets the standard of care. It also directs physicians to accredited CME on the applications, benefits, and risks of AI.[90] In May 2026 the FSMB extended this work by forming a Workgroup on the Regulation of Artificial Intelligence in the Practice of Medicine, charged with drafting recommendations or model guidelines for state boards, with particular attention to tools performing clinical functions under limited or no direct physician supervision.[91] The Royal College of Physicians and Surgeons of Canada is among the earliest national bodies to recommend embedding AI capability in an existing competency framework rather than in a stand-alone course, proposing that digital health literacy competencies be integrated into CanMEDS and that AI-related ethics be added as a resident sub-competency.[8] Both remain advisory: FSMB guidance informs but does not bind individual state boards, and no adoption or evaluation of the CanMEDS recommendation appears in the evidence reviewed here.

There is a growing set of openly accessible, clinician-facing offerings: the Society of Teachers of Family Medicine's Artificial Intelligence and Machine Learning for Primary Care (AiM-PC) curriculum, a free five-module program for medical students, residents, and practicing primary care clinicians carrying AAFP CME credit;[79] and the American College of Physicians' Generative AI for Internal Medicine Physicians primer, a brief (approximately 45–60 minutes), self-paced, web-based module provided to ACP members without CME or CC credit.[93]

Primary survey data collected for this analysis expose the operational tension behind this uneven provision (Section 5.3; Appendix I). The layer of guidance closest to practicing physicians is being produced by organizations that are themselves faculty-constrained, that weigh their members' greatest content need against competing educational priorities, and that remain uncertain which part of the curriculum is theirs to own.

Accreditation and certification bodies. Neither the LCME nor the ACGME currently holds a binding AI standard (though this is believed to be in progress), and AI is largely untested in licensing/board examinations; the clearest concrete board-level requirement is the ABR's 2021 addition of AI to its Qualifying (Core) Exam noninterpretive-skills section.[68] The Macy conference recommendations explicitly call on accrediting, licensing, and certifying bodies to incorporate AI competencies and to build a national best-practices network, documenting that this alignment is recommended but not yet enacted.[10] Interviews conducted for this analysis indicate that this reflects a timing problem as much as inaction, and point to anticipated graduate medical education standards whose timing is discussed in Section 4.5.

Two AAMC/MedBiquitous instruments provide organization-backed guidance: the IACAI AI Integration Framework provides parallel Learner and Educator Matrices across 12 domains and five role levels (intrapersonal, micro, meso, macro, mega), and a GME stakeholder/RACI mapping;[21] and the AAMC New and Emerging Areas in Medicine Competency Series: AI offers six cross-continuum domains across three developmental tiers.[20] These are guidance and draft instruments (not adopted accreditation standards) and are flagged as such throughout.

Multi-stakeholder coalition activity. The Coalition for Health AI (CHAI), a multi-stakeholder community of health systems, clinicians, developers, government, academia, and patient communities, has extended organization-level guidance to the practicing-clinician level. Its 2023 Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare set out consensus assurance principles and explicitly names trainees and continuing education among the stakeholder groups for trustworthy AI.[94] Its 2025–2026 activity adds a dedicated education workstream: an expert-led clinician webinar series; a seven-module Responsible AI in Nursing micro-credential developed with the Florida State University College of Nursing, scheduled to launch in spring 2026; high-level guidance on the responsible use of AI in health care co-published with the Joint Commission and intended to inform a planned voluntary certification program; and plans to develop an AI-literacy program for physicians with specialty societies, naming the AAFP and the College of American Pathologists as examples.[81] As with the accreditation-adjacent signals above, these remain guidance and provision rather than evaluated curricula; their significance here is as multi-stakeholder convening infrastructure that now reaches practicing clinicians, the level where formal educational structures are thinnest.

Cross-sector convening and consensus guidance. Three further convenings, outside the scoring corpus and used here only as advisory context, reinforce this picture. A 2026 NAM clinicians forum on AI in health and medicine advanced a Statement of Common Principles holding that clinical AI should be engaged, safe, effective, efficient, fair, accessible, accountable, transparent, secure, and adaptive, and framed adoption as the product of a tool's utility and clinicians' trust in it.[12] Recommendations from the 2025 Millennium Conference note that accountability (a clinician's professional standing, personal stake, and capacity to repair harm) is the dividing line between AI used as a tool and AI treated as a teammate, and that more capable models widen rather than close this accountability gap.[11] The 2023 National Academies workshop on AI in health professions education framed the field's twin tasks as training clinicians in AI and using AI in training (adding a third: defining the clinician's role in AI development and deployment), emphasized interprofessional education, and organized institutional action around Lomis and colleagues' eight steps[95]: educating faculty; building relationships; convening a local advisory group; revisiting competencies and curriculum; shifting assessment from possession of knowledge toward access, critical appraisal, and application; revisiting admissions; evaluating impact; and engaging nationally.[9] Like the society guidance above, all three are advisory and consensus-oriented rather than binding. The 2023 National Academies workshop's educational recommendations span the continuum: at the UME level, AI should be treated as a catalyst for higher-order competencies rather than a new stand-alone competency, protecting the intellectual struggle needed for deep learning while using AI to facilitate feedback and redesign assessment around metacognition and appraisal of AI outputs; at the GME level, foundational clinical skills should remain central because history taking and physical examination shape the prompts on which AI performance depends, and precision education uses of AI (e.g., linking encounters to targeted resources) should be introduced with attention to whose reasoning a documented note actually reflects; and across levels, faculty development must equip faculty to instruct on bias, error, and hallucination, use AI in probing questions and self-study assignments that permit productive failure, and support a growth mindset toward learning with AI. Assessment recommendations bracket this content: assessment of clinical competence should be conducted with AI available to the learner, some assessments should still be conducted without AI while high-stakes examinations remain AI-restricted, and high-stakes assessment should always include an element of human evaluation.

Guidance for clinical AI use by trainees and physicians. Society and accreditation guidance to date treats AI mainly as an educational competency rather than governing its day-to-day clinical use by learners and physicians. As an example, a 2026 review of five major medical and specialty organizations—the American Academy of Dermatology, the AMA, the ACGME, the AAMC, and the American Board of Medical Specialties—found that none had issued resident-specific guidance on AI-assisted clinical documentation; the AMA offered a general principle that AI be safe, accurate, and non-discriminatory, and the AAMC's education principles do not address resident documentation.[77] The same pattern holds for institutional policy: audits of US medical school AI documents find that formal, medicine-specific AI policies are the exception rather than the rule and concentrate on academic integrity rather than on the oversight and accountability structures needed for clinical use.[62,65] The result is a governance vacuum precisely where trainees and practicing physicians are beginning to use AI scribes and decision-support tools in patient care.

Where local governance does exist, primary data collected for this analysis indicate that it is often restrictive rather than enabling. The governance problem is therefore twofold: an absence of medicine-specific policy defining accountable clinical use, and a patchwork of localized platform restrictions, described in Sections 5.5 and 7.1, that determines what physicians can actually do with what they have been taught.

7. Cross-Cutting Gaps and Unifying Synthesis Across the Continuum

Across UME, GME, and CME, the landscape is conceptually mature and empirically immature at every level. The consolidated scores show a consistent shape: competency frameworks, formal curricula, and ethics/equity/legal sit at Developing; faculty development and validated assessment sit at Emerging; and the accreditation/licensing/society/CC bundle sits at Emerging at all three levels. No reporting domain at any level reaches Established in the consolidated view, and none reaches Evidenced.

7.1 Seven gaps that span the continuum

1. **No behavior or patient outcomes.** Across UME, GME, and CME, every reported outcome stops at reaction and knowledge (Kirkpatrick Levels 1–2); no study reaches behavior change (Level 3) or patient outcomes (Level 4). An independent systematic review of AI in health care and medical education, drawing on English- and Chinese-language databases, reached a parallel conclusion about the evidence itself, grading almost all 25 included studies at the two lowest levels of the Oxford Centre for Evidence-Based Medicine scale.[96]
2. **Absent validated assessment.** Validated AI-competency instruments are scarce at every level; the field relies on non-validated self-report, and the one recurring shared instrument (MAIRS-MS) is critiqued as mismatched to AI-literacy content.[48,55,56]
3. **Milestones/EPAs only proposed.** AI-specific milestones and EPAs exist only as proposals (e.g., emerging Core EPAs 14–16;[61] an ACGME-milestone-style radiology framework;[72] an EPA-for-AI concept[53]), none adopted or assessed.
4. **No binding accreditation or licensing.** No binding accreditation standard requires AI, and AI is largely absent from licensing/board examinations; the clearest concrete exception is a single specialty board's noninterpretive-skills addition.[68] At the licensing level, FSMB guidance recommends that physicians pursue accredited CME on AI, and its 2026 workgroup will draft model guidelines for state boards, but neither is binding and neither establishes a requirement that physicians demonstrate AI competency.[90,91]
5. **Faculty development named but unmet.** Faculty unreadiness is one of the most consistently documented barriers at every level, with no delivered, evaluated, financed faculty development program in the evidence. One effort is under development as this analysis is written: a commercial vendor has partnered with health professions educators on a 10-module web-based program to prepare faculty to use AI in their educational practice, though no outcome data are yet available. Primary data collected for this analysis rank it first among barriers at both the graduate and continuing levels: 50.0% of residency and fellowship program directors and 64.7% of society CPD directors named limited faculty expertise as a leading obstacle. The barrier is not new: an integrative review published in 2019 already identified the need for an experienced content specialist and the knowledge gap between physicians and engineers among the leading obstacles to implementing AI in medical education, and proposed collaboration with engineering and computing faculty as the practical remedy — the same remedy this report finds widely recommended but rarely formalized.[97]
6. **Reporting standards, dissemination, and scale-up are underdeveloped.** Standardized reporting for AI-in-medical-education scholarship is not established, hindering pooled learning across programs. Dissemination pathways for effective curricula and rapidly evolving practice guidance are similarly nascent; specialty journals focused on AI in medicine (e.g., NEJM AI), open-access medical-education repositories (e.g., MedEdPORTAL's AI collection), and organizational resource collections (e.g., the AAMC's Advancing

AI Across Academic Medicine resource collection) exist but are not integrated into a shared infrastructure that supports scale-up, particularly for lower-resourced institutions.

7. **Institutional and program-level AI-use governance is sparse and non-specific.** Beyond competency frameworks and accreditation, the policies and oversight structures that actually govern how students, residents, and practicing physicians may use AI in clinical work are underdeveloped. An audit of AI-related documents across US medical schools retrieved materials from 73.7% (146 of 198) of institutions but identified only 21 formal AI policies (10.8% of documents), with coverage dominated by academic-integrity and plagiarism content (81.5%) and audit, compliance, and infrastructure provisions each addressed by fewer than 7% of institutions.[65] A separate analysis found that 61% (121 of 199) of schools had any AI policy, most inherited from a broader university system rather than tailored to medical students, leaving roughly 40% with no explicit policy.[62] At the GME level, a review of five major medical and specialty organizations found no resident-specific guidance on AI-assisted clinical documentation, only general, non-binding principles.[77] Governance at the level of teaching hospitals and health systems, where trainees and physicians actually deploy these tools, remains largely unexamined in the reviewed literature. Primary data collected for this analysis extend the finding to practicing physicians: society CPD directors reported that hospital and health-system restrictions on which AI platforms physicians may use are a decisive determinant of what clinicians can do with what they have learned, so that without coherent local governance, national education efforts struggle to reach daily clinical workflows.

7.2 Where readiness drops across the UME→GME→CME transition

Implementation is most visible in UME, where literacy teaching shows knowledge gains and national curricular tracking documents rapid adoption, and in pockets of GME, overwhelmingly radiology, where a handful of evaluated residency curricula exist. It thins again at the practicing physician level: appetite is high (92% of physicians want more training[3]) but provision is heterogeneous, largely unevaluated, and concentrated in well-resourced settings. The articulation between levels is weak: competency proposals at one level are rarely carried forward as adopted, assessed expectations at the next, and the same gaps recur rather than resolving as learners progress. The AAMC New and Emerging Areas draft competency series is a partial exception in that it explicitly extends across UME, GME, and continuing practice as tiered developmental milestones, though adoption and assessment remain unproven. The clinical reality evidence reframes the central problem as one of timing (practice has moved ahead of education across the whole continuum, and the gap is widening in real time). Independent cross-sector convenings sharpen what is at stake in that gap. A 2026 NAM clinicians forum describes an emerging clinician–patient–AI triad, at risk of fragmenting into a misaligned tetrad when clinicians and patients rely on different tools, together with deskilling, mis-skilling, and never-skilling risks,[12] and the 2025 Millennium Conference noted the physician's enduring responsibility in accountable judgment that AI cannot hold.[11] The readiness target, on both accounts, is not tool operation but the clinician's accountable, interpretive role. This is also where the paradigms introduced in Section 1 converge in practice: the clinician being trained must appraise and remain accountable for AI within an LHS, apply it to PM use cases at the point of care, and be developed through PME that runs the same learning loop, reinforcing that readiness is participation in a learning system, not a discrete set of tool skills.[13,14,15] Trainee surveys now document this adoption–training inversion at each step of the continuum: near ubiquitous student LLM use with formal instruction reaching roughly one in ten,[29] emerging resident professional use with no reported formal training and near universal appetite for curriculum,[71] and broad physician adoption with thin training.[3]

7.3 Bottom line

Area	Current state	Biggest gap	Highest-priority action
UME	Broad but shallow; AI literacy the best-developed domain	Delivery inconsistent; assessment overwhelmingly formative and self-report	Longitudinal, spiral integration with validated assessment and AI-specific EPAs
GME	Early-stage and uneven; implementation concentrated in radiology	AI-specific ACGME milestones and EPAs are not in place	Advance milestone-style frameworks from proposal to in-force assessment
CME	Predominantly conceptual and pre-implementation; accredited provision growing	Clinical practice has already moved ahead of CME	Point-of-care, just-in-time formats prioritizing critical appraisal of real-world validity
Competency frameworks	Developing (2) at all three levels	No framework is standardized or accreditation-adopted	Adopt and assess a shared cross-continuum framework as a coordinating spine
Assessment and outcomes	Emerging (1) at all three levels	Outcomes stop at Kirkpatrick Levels 1–2; validated instruments scarce	Build and validate instruments and outcome measures reaching behavior and patient levels
Faculty development	Emerging (1) at all three levels; the most consistently named barrier	No delivered, evaluated, funded program in the evidence	Fund and deliver faculty development rather than only naming it
Accreditation, licensing, and CC	Emerging (1) at all three levels	No binding standard requires AI; AI largely absent from licensing and board examinations	Incorporate AI competencies into accreditation, licensing, certification, and CC
Institutional governance	Sparse and rarely specific to medicine or clinical practice	Policies emphasize academic integrity over clinical oversight and accountability	Issue medicine-specific policies governing clinical AI use by trainees and physicians
Dissemination and scale-up	Nascent; journals, repositories, and resource collections are not integrated	No shared infrastructure for scale-up, particularly for lower-resourced institutions	Adopt transparent reporting standards and build a national best-practices network

Current state reflects the consolidated six-domain scores in Sections 3.3, 4.4, and 5.4; gaps condense Section 7.1.

8. Strategic Recommendations, Priority Actions, and Integrated Competency Model

The recommendations below are based on a synthesis of available evidence, and the expertise and experience of the contributors to this report. They are framed as recommendations, not findings, and draw only on the evidence reviewed in this report, including the stakeholder interviews and surveys described in Appendix I. They align with the consensus action agenda in the Macy recommendations. They also converge with the 2026 NAM clinicians forum's Statement of Common Principles and with the 2025 Millennium Conference emphasis on accountable judgment and on protected, AI-free "sacred spaces" for skill-building and for assessment.

8.1 Recommendations by education level

Undergraduate Medical Education

- **Move from electives to longitudinal, spiral integration** of AI literacy, critical appraisal, ethics, and bias into existing organ-system blocks and clerkships, building on the current literacy base.
- **Adopt and assess a shared competency framework**, using the AAMC draft series and IACAI Matrices[20,21] as a coordinating spine.
- **Build validated assessment and pilot AI-specific EPAs**, advancing proposals such as emerging Core EPAs toward measured, behavior-level evaluation.

Graduate Medical Education

- **Develop AI-specific ACGME milestones/EPAs**, advancing milestone-style frameworks and ACGME-domain mappings from proposal to in-force assessment.
- **Make accountability for AI-assisted clinical work explicit and assessed**, rather than leaving it to individual supervisors. Program directors surveyed for this analysis described residents retaining ownership of the clinical note, and editing and checking AI-generated text for hallucination, as obligatory rather than discretionary. Because documentation and administrative tasks are the most common form of resident AI use (67.5%) and only 2.6% of program directors were very confident in residents' ability to appraise AI outputs, this expectation warrants statement in program policy and inclusion in assessment rather than being left implicit.
- **Establish governance for resident AI use**. As described in Section 4.5, new ACGME Common Program Requirements addressing AI are anticipated in September 2026 and would not take effect until 2028. That interval leaves sponsoring institutions and programs responsible for their own interim governance and usage policies during precisely the period in which resident AI use is becoming routine, so local policy development should not be deferred to a forthcoming national standard.

Continuing Medical Education

- **Prioritize critical appraisal of real-world validity** over benchmark performance, given the benchmark-to-bedside drop and physicians' trust in peer-reviewed/clinical evidence.
- **Meet physicians where they work** with hands-on, interprofessional, self-paced formats they prefer, and close the consumer-LLM exposure gap with institutionally validated tools.
- **Replace demand-driven planning with proactive needs assessment**. Nearly two-thirds of specialty society CPD directors surveyed for this analysis (64.7%) had not formally assessed member interest in AI content, and societies described an "unknown unknowns" problem in which clinicians cannot yet articulate what they need to learn. Needs assessment should therefore draw on observed practice patterns, quality and safety data, and institutional deployment plans rather than on unprompted member demand.
- **Shorten the production cycle and align content with institutional AI governance**. Societies reported that material risks obsolescence before or shortly after release because the development and review cycle is slower than the pace of tool change, and that members are frequently restricted by their hospitals or health systems as to which AI platforms they may use. Modular, rapidly updatable formats, and coordination with institutional governance so that the tools taught are tools learners may actually use, would address both constraints.

8.2 Recommendations for accreditors, professional societies, and policymakers

- **Incorporate AI competencies into accreditation, licensing, certification, and CC**: LCME/ACGME standards, United States Medical Licensing Examination (USMLE)/board content, and a CC AI pathway, building on the single existing board precedent.
- **Anchor physician AI readiness in the LHS, PM, and PME** so that training, clinical practice, and continuous improvement share one data-to-practice learning cycle: prepare physicians to appraise and remain accountable for AI within the LHS and PM workflows, and use PME to deliver the workflow-embedded, feedback-driven training that CME in particular will require.[13,14,15]
- **Harmonize fragmented society guidance** into shared, cross-specialty standards and a national best-practices network, reducing the specialty-siloed patchwork.
- **Fund ambitious, participatory scholarship** that generates the missing Level 3–4 outcome evidence and uses transparent reporting standards, partnering across the health professions and with industry where appropriate.
- **Address equity and resourcing** deliberately, since enabling conditions and training exposure concentrate in well-resourced systems and the evidence skews to North America and Europe.
- **Issue clear policy and best-practice guidance on the use of AI in creating educational content**. Continuing professional development directors surveyed for this analysis asked explicitly for such guidance, and no current accreditation standard or society statement addresses it, leaving individual content developers to set their own rules for disclosure, verification, and authorship of AI-assisted educational material.
- **Weigh the environmental impact of AI in guidance and governance**. Institutional and society guidance on responsible AI use should consider the energy, water, and emissions footprint of clinical AI alongside accuracy, cost, and safety, and vendor transparency about the resource use of deployed tools would give educators and institutions the information those judgments require.[16,17]

8.3 A proposed integrated competency model across the continuum

The evidence and proposed frameworks converge on a coherent, continuum-spanning model. We propose anchoring physician AI readiness in eight readiness domains: AI literacy; clinical application; critical appraisal; ethics/legal/governance; human–AI collaboration; professional identity/systems thinking; faculty readiness; and assessment strategies, expressed at three developmental tiers that mirror the AAMC draft series (Novice, Proficient, Expert).

Within that scaffold, the AAMC New and Emerging Areas domains (AI knowledge for practice; AI-assisted patient care and clinical reasoning; communicating about AI and interprofessional collaboration; professionalism and ethics; practice-based AI learning and upskilling; and systems-based AI practice)[20] provide the tier-by-tier behavioral anchors, and the IACAI Learner and Educator Matrices[21] provide the role-level actions (intrapersonal through mega) for learners and the educators who support them.

Three design principles follow directly from the cross-cutting gaps. First, every competency must carry a validated assessment and an outcome target reaching at least behavior change, not knowledge alone. Second, faculty readiness is a critical competency with its own delivered, financed development, not an assumed precondition. Third, the model must be adopted into accreditation, licensing, certification, and CC to become binding rather than advisory. The AAMC draft framework and IACAI matrices supply the coordinating spine; the missing pieces are validation, faculty capacity, outcome evidence, and accreditation uptake.

8.4 Priority areas for curriculum reform and faculty development

- **Curriculum:** longitudinal, spiral integration with assessment built in from the start; critical appraisal of real-world validity as a through-line; and continual updating as tools shift from classical machine learning (ML) to generative and agentic AI. Program directors and society education leaders alike favored interactive, case-based, and workshop formats over didactics, and ranked recognizing the limitations of AI and appraising its outputs above tool operation as content priorities.
- **Faculty development:** delivered, evaluated, funded programs and protected time, with interdisciplinary co-teaching formalized into reproducible structures rather than one-off projects. Faculty capacity was the single most frequently named barrier in both surveys conducted for this analysis, cited by 50.0% of program

directors and 64.7% of society continuing professional development directors, and program directors described it as a precondition for teaching or assessing safe AI use rather than as an accompaniment to it.

- **Roles and interprofessional preparation:** As AI is integrated into clinical settings, the roles, responsibilities, and accountability of physicians, and of the other health professionals they work with, will shift; curricula at every level should therefore include interprofessional education that rehearses team-based decision-making, appropriate task delegation, and accountability with AI in the loop, expanding on the isolated interprofessional examples documented above into standard practice. Interviews conducted for this analysis add a placement caveat: interprofessional AI learning is currently concentrated in undergraduate medical education rather than in the active clinical care environments where AI-enabled teams operate, so it should extend into workplace-based training and continuing education rather than remaining a preclinical exercise.

8.5 Conclusion

Physician AI-readiness education is advancing across the continuum but remains early-stage: strong on rationale, frameworks, and literacy, and weak on validated assessment, faculty capacity, outcome evidence, and binding accreditation. The same seven gaps recur at UME, GME, and CME, and readiness drops precisely where appetite is highest, at the practicing-physician level, even as clinical practice moves ahead of education.

The path forward is to convert proposals into adopted, assessed, and funded practice: validating competencies, generating behavior- and patient-level outcomes, building faculty capacity, and aligning accreditation, licensing, certification, and CC. The stakeholder surveys and interviews reported in Appendix I sharpen these priorities; the evidence reviewed here already makes their direction clear.

Appendix A. References

References are numbered in order of first appearance in this report.

1. Boscardin CK, Abdulnour R-EE, Gin BC. Macy Foundation Innovation Report Part I: current landscape of artificial intelligence in medical education. Acad Med. 2025;100(9 Suppl):S15-S21. doi:10.1097/ACM.0000000000006107.

2. Ejaz H, McGrath H, Wong BLH, Guise A, Vercauteren T, Shapey J. Artificial intelligence and medical education: a global mixed-methods study of medical students' perspectives. Digit Health. 2022;8:1-11.

3. American Medical Association, Center for Digital Health and AI. 2026 Physician Survey on Augmented Intelligence. March 2026. [Stakeholder survey; n=1,692; fielded Jan 15–Feb 2, 2026.]

4. American Medical Association, Center for Digital Health and AI, with the Digital Medicine Society (DiMe). From Benchmark to Bedside: The Current State of AI in Clinical Practice. Landscape Report, May 2026.

5. Gong EJ, Bang CS, Lee JJ, Baik GH. Knowledge-practice performance gap in clinical large language models: systematic review of 39 benchmarks. J Med Internet Res. 2025;27:e84120. doi:10.2196/84120.

6. Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. JAMA. 2025;333(4):319-328. doi:10.1001/jama.2024.21700.

7. Ng FYC, Thirunavukarasu AJ, Cheng H, et al. Artificial intelligence education: an evidence-based medicine approach for consumers, translators, and developers. Cell Rep Med. 2023;4(10):101230. doi:10.1016/j.xcrm.2023.101230.

8. Royal College of Physicians and Surgeons of Canada. Task Force Report on Artificial Intelligence and Emerging Digital Technologies. Reznick RK, Harris K, Horsley T, Sheikh Hassani M, prepared by; in collaboration with the Council Task Force on Artificial Intelligence and Emerging Digital Technologies. Ottawa, ON: Royal College of Physicians and Surgeons of Canada; February 2020.

9. National Academies of Sciences, Engineering, and Medicine. Artificial Intelligence in Health Professions Education: Proceedings of a Workshop. Cuff PA, Hammers Forstag E, rapporteurs. Washington, DC: The National Academies Press; 2023. doi:10.17226/27174.

10. Conference participants. Josiah Macy Jr. Foundation Conference on Artificial Intelligence in Medical Education: proceedings and recommendations. Acad Med. 2025;100(9 Suppl):S4-S14. doi:10.1097/ACM.0000000000006099.

11. Schwartzstein R, Dubosh N, LaConte L, Rodman A, Samuel A, Zheng J; Shapiro Institute for Education and Research, Beth Israel Deaconess Medical Center/Harvard Medical School. Artificial intelligence: prompts, hallucinations and the future of medical education. Recommendations from the Millennium Conference 2025, presented at: AAMC annual meeting; 2025. https://www.shapiroinstitute.org/mc2025. Accessed July 4, 2026.

12. National Academy of Medicine. Clinicians Forum on Artificial Intelligence in Health & Medicine. Washington, DC: National Academy of Medicine; April 2, 2026. https://nam.edu/our-work/programs/leadership-consortium/evidence-digital-health/clinicians-forum-on-ai-in-health-medicine/.

13. Dzau VJ, Ginsburg GS, Chopra A, Goldman D, Green ED, Leonard DGB, McClellan M, Plump A, Terry SF, Yamamoto KR. Realizing the full potential of precision medicine in health and health care: a vital direction for health and health care. NAM Perspect. 2016;6(9):1-13. doi:10.31478/201609k.

14. Steel PAD, Wardi G, Harrington RA, Longhurst CA. Learning health system strategies in the AI era. npj Health Syst. 2025;2:21. doi:10.1038/s44401-025-00029-0.

15. Desai SV, Khan S, Lomis KD. AI-enabled precision-education systems — transforming lifelong learning in medicine. N Engl J Med. 2026;394(9):838-841. doi:10.1056/NEJMp2512935.

16. Verma S, Philipsborn R, James C, Sullivan GM. A call to address the generative AI-environment paradox in graduate medical education. J Grad Med Educ. 2026;18(2):97-101. doi:10.4300/JGME-D-26-00169.1.

17. Ramachandran A, Sarabu C, Gupta U, Ghose S, Lee VS. Sustainably advancing health AI: a decision framework to mitigate the energy, emissions, and cost of AI implementation. NEJM Catal Innov Care Deliv. 2025;6(10). doi:10.1056/CAT.25.0125.

18. Lindqwister AL, Hassanpour S, Lewis PJ, Sin JM. AI-RADS: an artificial intelligence curriculum for residents. Acad Radiol. 2021;28(12):1810-1816.

19. Wong A, Sussman J, Price N, Makar M, Li B, Yang J, James C. The Data-Augmented, Technology-Assisted Medical Decision Making (DATA-MD) curriculum. Acad Med. 2025;100(9):1035-1039.

20. Association of American Medical Colleges. AAMC New and Emerging Areas in Medicine Competency Series: Artificial Intelligence. Draft. Washington, DC: AAMC; 2026. https://www.aamc.org/media/84666/download?attachment.

21. Association of American Medical Colleges, MedBiquitous; in collaboration with AMEE, AAHCI, IAMSE, APMEN. IACAI AI Integration Framework: Learner Matrix, Educator Matrix, and Graduate Medical Education Stakeholder and RACI Mapping (v2026). International Advisory Committee for Artificial Intelligence; 2026. https://www.medbiq.org/initiatives/international-advisory-committee-artificial-intelligence.

22. Gin BC, LaForge K, Burk-Rafel J, Boscardin CK. Macy Foundation Innovation Report Part II: from hype to reality: innovators' visions for navigating AI integration challenges in medical education. Acad Med. 2025;100(9 Suppl):S22-S29. doi:10.1097/ACM.0000000000006117.

23. Aydin Z, Chen JH, et al. Artificial Intelligence in Clinical Practice: A Curriculum for Medical Students. Stanford University; accessed via provided course outline. https://med.stanford.edu/content/dam/sm/ai-in-meded/documents/AI%20Curriculum--Module%20List.pdf. [Evolving curriculum document; learning objectives mapped to Delphi competency studies.]

24. Chang L, Sreedhar R. Integrating AI literacy into medical education: preparing future clinicians for an AI-driven healthcare system. Med Sci Educ. 2026;36:17-22.

25. Tan JA, Ng IKS, Hong A, Teo DB, Tan LF. Empowering the next generation: integrating artificial intelligence education into medical training. Postgrad Med J. 2026;102(1204):185-188.

26. Ötleş E, James CA, Lomis KD, Woolliscroft JO. Teaching artificial intelligence as a fundamental toolset of medicine. Cell Rep Med. 2022;3:100824.

27. Greengrass C. Navigating the AI revolution in medicine—adopting strategies for medical education. Med Sci Educ. 2025;35:1055-1061.

28. Patel MS, Dee EC, Toby S, Ramos M, Celi LA. The physician as a catalyst: reimagining medical education for an AI era. Med Teach. 2026. doi:10.1080/0142159X.2026.2653205.

29. Barr AA, Rozman RC, Liu K, et al. Medical students' use of large language models: a national survey. Int J Med Inform. 2026;220:106589. doi:10.1016/j.ijmedinf.2026.106589.

30. McGee RG. The digital native paradox: a framework for critical appraisal of generative AI in paediatric medical education. J Paediatr Child Health. 2026 (advance online; jpc.70386).

31. Feigerlova E. How to teach generative artificial intelligence in undergraduate medical education. Clin Teach. 2026;23:e70420.

32. Abdulnour R-EE, Gin B, Boscardin CK. Educational strategies for clinical supervision of artificial intelligence use. N Engl J Med. 2025;393(8):786-797.

33. Gordon M, Daniel M, Ajiboye A, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach. 2024;46(4):446-470.

34. Chan SCC, Young H, Parekh R. Medical students' and faculty members' perceptions and experiences of AI integration in health care practice and in medical curricula: a meta-ethnographic review. Med Educ. 2026;60(5):492-504. doi:10.1111/medu.70071.

35. Krive J, Isola M, Chang L, Patel T, Anderson M, Sreedhar R. Grounded in reality: artificial intelligence in medical education. JAMIA Open. 2023;6(2):ooad037.

36. Agarwal G, Ramamoorthi L, Yuen T, Merced E, Brenner J, Wu W, Sabina R. Exploring applications of artificial intelligence tools in clinical care and health professions education: an online module for students. MedEdPORTAL. 2025;21:11524.

37. Pizzolla I, Aro R, Duez P, De Lièvre B, Briganti G. Integrating artificial intelligence into medical education: lessons learned from a Belgian initiative. J Interact Learn Res. 2023;34(2):401-424.

38. Mihalas GI, Cotoraci C, Boru C. Integrating artificial intelligence in the medical curriculum: approaches and strategies. Stud Health Technol Inform. 2025. doi:10.3233/SHTI251069.

39. Rezazadeh H, Madadi Mahani A, Salajegheh M. Preparing Generation Z of health professions for artificial intelligence revolution through hacking education: an interventional study. Health Sci Rep. 2025;8:e71169.

40. Yao MS, Huang L, Leventhal E, Sun C, Stephen SJ, Liou L. Leveraging datathons to teach AI in undergraduate medical education: case study. JMIR Med Educ. 2025;11:e63602. doi:10.2196/63602.

41. Ahsan Z. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. BMC Med Educ. 2025;25:1187.

42. Rincón EHH, Jimenez D, Chavarro Aguilar LA, Pérez Flórez JM, Romero Tapia ÁE, Jaimes Peñuela CL. Mapping the use of artificial intelligence in medical education: a scoping review. BMC Med Educ. 2025;25:526. doi:10.1186/s12909-025-07089-8.

43. Komasawa N, Yokohira M. Learner-centered experience-based medical education in an AI-driven society: a literature review. Cureus. 2023;15(10):e46883.

44. Komasawa N, Yokohira M. Generative artificial intelligence (AI) in medical education: a narrative review of the challenges and possibilities for future professionalism. Cureus. 2025;17(6):e86316.

45. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med. 2021;96(11S):S62-S70. doi:10.1097/ACM.0000000000004291.

46. Farmakidis AL, Singh A, Leaf K, Rossi A. Artificial intelligence curricula in U.S. and Canadian medical schools. Data Snapshot. Washington, DC: Association of American Medical Colleges (AAMC); 2025.

47. Liaw W, Kueper JK, Lin S, Bazemore A, Kakadiaris I. Competencies for the use of artificial intelligence in primary care. Ann Fam Med. 2022;20(6):559-563. doi:10.1370/afm.2887.

48. Doherty G, McLaughlin L, Hughes C, McConnell J, Bond R, McFadden S. A scoping review of educational programmes on artificial intelligence (AI) available to medical imaging staff. Radiography. 2024;30:474-482.

49. Weightman MJ, Chur-Hansen A, Clark SR. AI in psychiatric education and training from 2016 to 2024: scoping review of trends. JMIR Med Educ. 2025;11:e81517.

50. Meade SM, Salas-Vega S, Nagy MR, Sundar SJ, Steinmetz MP, Benzel EC, Habboub G. A pilot remote curriculum to enhance resident and medical student understanding of machine learning in healthcare. World Neurosurg. 2023;180:e142-e148.

51. Hunt VM, Sprehe LK, de Lomba WC, et al. What the AI era doctor should know: a scoping review of proposed artificial intelligence competencies for medical education. npj Digit Med. Published online 2026. doi:10.1038/s41746-026-02761-9.

52. Russell RG, Novak LL, Patel M, Garvey KV, Craig KJT, Jackson GP, Moore D, Miller BM. Competencies for the use of artificial intelligence–based tools by health care professionals. Acad Med. 2023;98(3):348-356. doi:10.1097/ACM.0000000000004963.

53. McGrath SP, Kim KK, Johl K, Wang H, Anderson N. AI-PACE: a framework for integrating AI into medical education. npj Digit Med. Published online 2026. doi:10.1038/s41746-026-02768-2.

54. Park SH, Pinto-Powell R, Thesen T, et al. Preparing healthcare leaders of the digital age with an integrative AI curriculum: a pilot study. Med Educ Online. 2024;29:2315684.

55. Tolentino R, Baradaran A, Gore G, Pluye P, Abbasgholizadeh-Rahimi S. Curriculum frameworks and educational programs in AI for medical students, residents, and practicing physicians: scoping review. JMIR Med Educ. 2024;10:e54793.

56. Cheng D, Chan EYS, Song Y, Cui A, Chang A, Wang W. Framework for AI integration into the undergraduate medical curricula: a scoping review. BMC Med Educ. 2026;26:465.

57. Wood EA, Ange BL, Miller DD. Are we ready to integrate artificial intelligence literacy into medical school curriculum: students and faculty survey. J Med Educ Curric Dev. 2021;8:1-5.

58. Blanco MA, Nelson SW, Ramesh S, Callahan CE, Josephs KA, Jacque B, Baecher-Lind LE. Integrating artificial intelligence into medical education: a roadmap informed by a survey of faculty and students. Med Educ Online. 2025;30(1):2531177.

59. Charow R, Jeyakumar T, Younus S, et al. Artificial intelligence education programs for health care professionals: scoping review. JMIR Med Educ. 2021;7(4):e31043.

60. D'Souza RF, Surapaneni KM, Sathyanarayanan P, et al. Convergence of diverse expertise: a multidisciplinary training on the ethics of artificial intelligence in healthcare technology and research. J Acad Ethics. 2025;23:885-899.

61. Jacobs SM, Lundy NN, Issenberg SB, Chandran L. Reimagining core EPAs for UME in the era of AI. JMIR Med Educ. 2023;9:e50903.

62. Knopf E, Mohamed AA, Leidy L, Perumareddi P, Luck GR. Bridging the AI policy gap in medical education: assessing the lack of standardised guidelines in US medical schools. Clin Teach. 2026;23(2):e70347. doi:10.1111/tct.70347.

63. James CA, Wheelock KM, Woolliscroft JO. Machine learning: the next paradigm shift in medical education. Acad Med. 2021;96(7):954-957.

64. Crotty E, Singh A, Neligan N, Chamunyonga C, Edwards C. Artificial intelligence in medical imaging education: recommendations for undergraduate curriculum development. Radiography. 2024;30(Suppl 2):S67-S73. doi:10.1016/j.radi.2024.10.008.

65. Rush E, Byram JN, Garnett CN, et al. An audit of AI-related documents across US medical schools: a framework-based qualitative content analysis. Med Teach. 2026;48(3):493-505. doi:10.1080/0142159X.2025.2564869.

66. McFarlane A, Sadri S, Schwartz E, Gowda D. Resident physicians' perceptions of artificial intelligence and implications for medical education: a qualitative study. MedEdPublish. 2025;15:29.

67. Yabes JM Jr, Lindholm DA, Ford MB, Kiley J, Geringer JL. Artificial intelligence in military graduate medical education: trainee perceptions and current use. Mil Med. 2026 (advance article). doi:10.1093/milmed/usag099.

68. Garin SP, Zhang V, Jeudy J, Parekh VS, Yi PH. Systematic review of radiology residency artificial intelligence curricula: preparing future radiologists for the AI era. J Am Coll Radiol. 2023;20(6):561-569.

69. Talwalkar JS, Rafi W, Loza A. Artificial Intelligence in Clinical Care. Yale Primary Care Pediatrics Curriculum. Yale School of Medicine Department of Pediatrics; 2026. https://pcpc.yale.edu. Accessed July 5, 2026.

70. Nadarajah A, Malekzadeh G, Ratnapalan S. The use of artificial intelligence in paediatric postgraduate medical education: a scoping review. Paediatr Child Health. 2025;30(8):694-704.

71. Fried AJ, Dorn SD, Leland WJ, Mullen E, Williams DM, Zaas AK, MacGuire J, Bynum DL. Large language models in internal medicine residency: current use and attitudes among internal medicine residents. Discov Artif Intell. 2024;4:70. doi:10.1007/s44163-024-00173-w.

72. Jassar S, Zhou Z, Leonard S, Youssef A, Probyn L, Kulasegaram K, Adams SJ. Educational competencies for artificial intelligence in radiology: a scoping review. Acad Radiol. 2025;32(9):5624-5634.

73. Tolentino R, Hersson-Edery F, Yaffe M, Abbasgholizadeh-Rahimi S. AIFM-ed curriculum framework for postgraduate family medicine education on artificial intelligence: mixed methods study. JMIR Med Educ. 2025;11:e66828.

74. Abernethy J, Shah A, Chen B, Reynolds S, Wright SM, O'Rourke P. Integrating AI scribes into medical education: guardrails for preserving clinical reasoning. J Gen Intern Med. 2026;41(9):2598-2602.

75. Zheng H, Zhang X. Psychiatry in the age of AI: transforming theory, practice, and medical education. Front Public Health. 2025;13:1660448.

76. Mori Y, Kopylov U, Sinonquel P, et al. Curriculum for safe and effective use of artificial intelligence in endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy. 2026;58:202-210. doi:10.1055/a-2742-4342.

77. Giordano JA, Jones E. AI scribe use in residency training: a call for specialty society guidance in graduate medical education. Adv Med Educ Pract. 2026;17:578656. doi:10.2147/AMEP.S578656.

78. Velez-Florez MC, Ghosh A, Patton D, Sze R, Reid JR, Sotardi S. Artificial intelligence curriculum needs assessment for a pediatric radiology fellowship program: what, how, and why? Acad Radiol. 2023;30(2):349-358.

79. Liaw W, Hischier B, James CA, Kakadiaris I, Kueper JK, Rahimzadeh V. Bridging the gap: transforming primary care through the artificial intelligence and machine learning for primary care (AiM-PC) curriculum. Ann Fam Med. 2024;22(6):570-571. doi:10.1370/afm.240537.

80. American Board of Artificial Intelligence in Medicine. ABAIM AI Courses. https://abaim.org/courses. Accessed July 2026.

81. Coalition for Health AI. 2025 Impact Report: Showcasing Key Milestones and Member Community. Coalition for Health AI; March 11, 2026. https://www.chai.org/news/chai-releases-2025-impact-report-showcasing-key-milestones-and-member.

82. Dumić-Čule I, Orešković T, Brkljačić B, Kujundžić Tiljak M, Orešković S. The importance of introducing artificial intelligence to the medical curriculum—assessing practitioners' perspectives. Croat Med J. 2020;61:457-464.

83. Sacchi de Camargo Correia G, Hazim AZ, Zheng-Lin B, Gannon N, Boyle AWR, Crowe S, Jones J, Manochakian R. Study of the perceptions and concerns of a single health system hematology and oncology workforce about AI in clinical practice and medical education. JCO Clin Cancer Inform. 2026;10:e2500296. doi:10.1200/CCI-25-00296.

84. Mehta N, Mehta S, Rubenstein A, Wood SK. Not replaced, but reinvented: AI education pathways to prepare future physicians to lead healthcare transformation. Perspect Med Educ. 2025;14(1):849-859.

85. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ. 2019;5(2):e16048.

86. American Medical Association. AMA Principles for Augmented Intelligence Development, Deployment, and Use. Chicago, IL: American Medical Association; November 14, 2023. https://www.ama-assn.org/system/files/ama-ai-principles.pdf.

87. Association of American Medical Colleges. Principles for the Responsible Use of Artificial Intelligence in and for Medical Education. Version 2.0. Washington, DC: AAMC; July 31, 2025. https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use.

88. National Academy of Medicine; The Learning Health System Series; Krishnan S, Matheny M, Fontaine E, et al., eds. An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action. Washington, DC: National Academies Press; 2025. doi:10.17226/29087.

89. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models. Geneva: World Health Organization; January 18, 2024.

90. Federation of State Medical Boards. Navigating the Responsible and Ethical Incorporation of Artificial Intelligence into Clinical Practice. Adopted by the FSMB House of Delegates, April 2024. Euless, TX: Federation of State Medical Boards; 2024.

91. Federation of State Medical Boards. Federation of State Medical Boards announces new workgroup on regulation of artificial intelligence in medical practice [news release]. Euless, TX: FSMB; May 19, 2026. https://www.fsmb.org/advocacy/news-releases/fsmb-announces-new-workgroup-on-regulation-of-artificial-intelligence-in-medical-practice/.

92. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med. 2024;99(1):22-27.

93. American College of Physicians. Generative AI for Internal Medicine Physicians [online module]. American College of Physicians; developed with Adam Rodman, MD (Harvard Medical School). https://assets.acponline.org/generative-ai-module/. Accessed July 5, 2026.

94. Coalition for Health AI. Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare. Version 1.0. Coalition for Health AI; April 4, 2023. https://chai.org.

95. Lomis K, Jeffries P, Palatta A, Sage M, Sheikh J, Sheperis C, Whelan A. Artificial intelligence for health professions educators. NAM Perspect. Discussion paper. Washington, DC: National Academy of Medicine; 2021. doi:10.31478/202109a.

96. Sun L, Yin C, Xu Q, Zhao W. Artificial intelligence for healthcare and medical education: a systematic review. Am J Transl Res. 2023;15(7):4820-4828.

97. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ. 2019;5(1):e13930. doi:10.2196/13930.

98. Lee YM, Kim S, Lee YH, Kim HS, Seo SW, Kim H, Kim KJ. Defining medical AI competencies for medical school graduates: outcomes of a Delphi survey and medical student/educator questionnaire of South Korean medical schools. Acad Med. 2024;99(5):524-533.

99. Association of American Medical Colleges; American Association of Colleges of Osteopathic Medicine. Curriculum SCOPE (Structure, Content, Organization, Process, Evaluation) Survey: artificial intelligence in the medical school curriculum, 2022–2023 to 2024–2025. Washington, DC: AAMC/AACOM; 2026. https://www.aamc.org/SCOPE.

100. Triola MM, Rodman A. Integrating generative artificial intelligence into medical education: curriculum, policy, and governance strategies. Acad Med. 2025;100(4):413-418.

101. Mendes JM. Reimagining healthcare education through nurturing AI-driven innovation. BMC Med Educ. 2025;25:1644.

102. Boman M, Jhun P, Schaekermann M. Scaffolding for success: blending learning with and about generative AI in medical education. Med Teach. 2025;47(12):1911-1917.

103. James CA, Wachter RM, Woolliscroft JO. Preparing clinicians for a clinical world influenced by artificial intelligence. JAMA. 2022;327(14):1333-1334.

104. Waldman CE, Hermel M, Hermel JA, et al. Artificial intelligence in healthcare: a primer for medical education in radiomics. Per Med. 2022;19(5):445-456.

Appendix B. Evidence Scale, Source Codes, and Readiness Domains

Field-level evidence scale (0–4)

Score	Label	Definition
0	**Absent**	No evidence in the literature — assigned only where the evidence indicates absence, not merely because a topic is unmentioned.
1	**Emerging**	Isolated examples; described in commentary or opinion; no systematic implementation.
2	**Developing**	Multiple documented examples; some frameworks proposed; inconsistent across institutions/programs.
3	**Established**	Widespread adoption; supported by professional-society guidance or accreditation standards.
4	**Evidenced**	Systematic implementation with outcome data demonstrating effectiveness.

Source codes & evidence-quality levels

Code / Level	Definition
L	Peer-reviewed literature
P	Policy or accreditation document
G	Gray literature, professional-society guideline, or white paper
O	Other, including primary stakeholder survey and interview data collected for this analysis (see Appendix I)
Quality: High	Strong empirical design (RCT, prospective cohort, large survey), n > 50, validated instruments, outcome data reported.
Quality: Moderate	Mixed methods, case series, small survey (n = 10–50), some measurement validity, limited outcome data.
Quality: Low	Expert commentary, opinion piece, single case study, no outcome data, no control group.

Eight readiness domains

Domain	Definition
AI Literacy	Foundational understanding of how AI/ML systems work — basic concepts, limitations, terminology.
Clinical Application	Applying AI tools in direct patient care, including clinical decision support.
Critical Appraisal	Evaluating AI tools for validity, bias, evidence quality, and contextual appropriateness.
Ethics / Legal / Governance	Ethical, legal, and policy frameworks governing AI use, and the specific institutional and program-level policies, oversight structures, accountability, and consent practices implemented (or, where absent, identified as gaps).
Human–AI Collaboration	Working alongside AI, including when to override or defer to AI recommendations.
Professional Identity / Systems Thinking	How AI integration shapes professional roles, responsibilities, and health-system dynamics.
Faculty Readiness	Preparation and support of educators to teach AI competencies, including faculty development.
Assessment Strategies	Methods to measure AI-related knowledge, skills, and attitudes in learners.

Crosswalk: reporting domains and readiness domains

The six reporting domains describe how the evidence base is organized and scored; the eight readiness domains describe what physicians need to be able to do. The two sets are related but not interchangeable, and the crosswalk below shows where they meet.

Reporting domain (Sections 3.3, 4.4, 5.4)	Sub-domains scored in Appendix C	Readiness domains primarily evidenced
1. Defined competency frameworks	Defined competency frameworks	Specifies competency content across domains. The convergent common core reported in Section 2 spans AI literacy, clinical application, critical appraisal, ethics/legal/governance, human–AI collaboration, and professional identity.
2. Formal curricula	AI literacy; clinical application (incl. clinical decision support); critical appraisal of AI tools; specialty-specific	AI literacy; clinical application; critical appraisal
3. Faculty development	Qualified faculty availability; interdisciplinary collaboration; budget / resources allocated	Faculty readiness
4. Validated assessment methods	Validated methods / instruments; milestones or EPAs; summative vs. formative; outcomes (e.g., patient) data	Assessment strategies
5. Ethics / equity / legal	Policy and governance; algorithmic bias	Ethics / legal / governance
6. Accreditation / licensing / board exam / society guidance / CC	Accreditation standards requiring AI; licensing / board exam AI content; professional society guidance; CC / AI certification pathway	None uniquely. This domain is the mechanism through which competencies in any readiness domain could become binding rather than advisory.

Two of the eight readiness domains carry no dedicated sub-domain in the scored matrices. Human–AI collaboration recurs as a cross-cutting theme in all three level-specific narratives (Sections 3.1, 4.1, and 5.1); professional identity / systems thinking is framed as an educational objective in Section 1 and appears in the UME and CME narratives (Sections 3.1 and 5.1). Both are retained in the proposed integrated competency model (Section 8.3), but neither is separately scored, so the six-domain and 18-sub-domain matrices should not be read as measuring them.

Appendix C. Detailed 18-Sub-Domain Matrices by Level

The three detailed 18-sub-domain matrices that underlie the consolidated six-domain scores in Sections 3.3, 4.4, and 5.4 are reproduced below.

Detailed 18-sub-domain matrix — UME

Sub-domain	Score (codes)	Supporting evidence & synthesis
A. Defined Competency Frameworks		
Defined Competency Frameworks	**2 (L,P,G)**	Frameworks converge on a common core of competencies; Lee YM 2024 is psychometrically validated (CFI .919, RMSEA .059);[98] Russell, Liaw, AI-PACE, Cheng, and Hunt add others, but none is standardized or accreditation-adopted.[47,51,52,53,56] **Developing**.
B. Formal Curricula		
AI Literacy	**2 (L,G)**	The most consistently implemented domain and the only one with breadth-and-outcomes: delivered courses report significant short-term gains (Krive 61%→97%; Agarwal 74%→87%, P<.001)[35,36,39,54] and AAMC data show schools with AI in the curriculum rising from 88 of 166 to 140 of 182 responding schools (53%→77%) in one year,[46] and the AAMC/AACOM Curriculum SCOPE Survey extends this trajectory to 83% of schools reporting AI in the curriculum by 2024–2025, with 90% of those schools (139) placing AI in the required curriculum. These growth figures describe the presence of AI content, not its depth or breadth: within the required curriculum, AAMC reports the leading topics as basic AI concepts (73%) and social, ethical, and legal aspects (70%), with AI in care delivery (54%), AI for learning and enrichment (44%), data analysis and evaluation (39%), and research and development (30%) covered less often, and the corresponding hours, sequencing, and assessment strategies are not captured.[99] **Developing**
Clinical Application (incl. clinical decision support)	**2 (L)**	Decision support, predictive analytics, and EHR-embedded tools are widely described and partially operationalized (AAMC reports AI in care delivery within the required curriculum in 54% of schools reporting AI content), but effectiveness is measured as knowledge, not clinical performance.[46,55,56] **Developing**.
Critical Appraisal of AI Tools	**2 (L)**	A near-universal objective with proposed scaffolds (VALIDATE and VERIFY frameworks),[30,31] yet addressed in only ~37% of program papers and rarely assessed.[55] **Developing.**
Specialty-Specific	**2 (L)**	Content exists across radiology/imaging (dominant), psychiatry, neurosurgery, and primary care but is imaging-concentrated and thin within specialties, with a recurrent generalist gap.[47,48,49,50] **Developing**.
C. Faculty Development		
Qualified Faculty Availability	**1 (L)**	A consistently documented barrier (Wood: 36% of faculty lacked basic AI understanding; faculty self-identify as novices); isolated, unevaluated faculty sessions exist.[57,58,59] **Emerging**.
Interdisciplinary Collaboration	**2 (L)**	Widely endorsed and partially demonstrated (six-discipline ethics workshop; multidisciplinary teams across medicine, data science, engineering, informatics, and ethics),[36,52,60] but project-specific, not standardized. **Developing.**
Budget / Resources Allocated	**1 (L,G)**	Addressed chiefly as a barrier; isolated cost models (≈US$28/participant datathon) and AAMC infrastructure gaps (30% secure AI access; 47% appropriate-use policy).[40,46] **Emerging**.
D. Assessment		

Sub-domain	Score (codes)	Supporting evidence & synthesis
Validated Methods / Instruments	**1 (L)**	Largely aspirational; the adapted MAIRS-MS recurs but is critiqued as mismatched to AI-literacy content, and most studies use non-validated self-report.[48,55,56] **Emerging**.
Milestones or EPAs	**1 (L,P)**	Conceptual only; Jacobs proposes emerging EPAs 14–16 against the AAMC Core EPA framework and Russell maps to ACGME levels, but none is adopted or assessed for AI.[47,52,61] **Emerging**.
Summative vs. Formative	**2 (L)**	Formative assessment is multiply documented (pre/post quizzes, reflections, pass/fail), with a few summative components (graded exams/reports), but the distinction is inconsistent.[35,36,37,38] **Developing**.
Outcomes (e.g., patient) Data	**2 (L)**	Short-term educational outcomes are multiply documented; no study reaches Kirkpatrick Level 3 (behavior) or 4 (patient), an absence repeatedly flagged. Patient-outcome data are not addressed.[35,51,59] **Developing**; limited by absent higher-order outcomes.
E. Ethics / Equity / Legal		
Policy & Governance	**2 (L,P,G)**	Among the best-represented areas, society/policy-anchored (AMA, AAMC, NAM, WHO, GDPR/EU AI Act); institutional adoption is growing but immature (61% of US schools have a policy, mostly generic; 47% have an appropriate-use policy).[46,62,100] **Developing**.
Algorithmic Bias	**2 (L)**	Robustly and consistently addressed as core content (unrepresentative data, Western-trained models, named fairness tools), but rarely embedded in evaluated curricula or assessed.[47,52,75] **Developing**.
F. Accreditation, Licensing & Professional Guidance		
Accreditation Standards Requiring AI	**1 (L,P,G)**	No binding standard mandates AI in general UME; the strongest signals are specialty/soft guidance (UK Health and Care Professions Council [HCPC] radiographer AI awareness; AAMC Principles).[46,48,63] **Emerging**.
Licensing / Board Exam AI Content	**1 (L,P)**	Documented as essentially absent (AI remains absent from the USMLE); only proposals to add AI to the MCAT/USMLE exist.[34,54,85] **Emerging**.
Professional Society Guidance	**2 (L,P,G)**	Abundant and widely cited (AMA, AAMC, NAM, WHO, Royal College of Canada, FSMB, Topol, multiple imaging bodies) but fragmented, specialty-siloed, and non-binding.[47,48,64,92] **Developing**.
CC / AI Certification Pathway	**1 (L,G)**	No formal CC/AI pathway operates; isolated exemplars (ABAIM certification; STFM AiM-PC; Harvard/Stanford/Coursera/AMIA certificates) are illustrations, not a pathway.[47,84,101] **Emerging**.

Detailed 18-sub-domain matrix — GME

Sub-domain	Score (codes)	Supporting evidence & synthesis
A. Defined Competency Frameworks		
Defined Competency Frameworks	**2 (L,G)**	Multiple GME-relevant frameworks are proposed: an ACGME-milestone-style radiology framework, an AIFM-ed family-medicine framework mapped to CanMEDS-FM, ACGME-mapped subcompetencies, and a Miller's-Pyramid adaptation, but all remain conceptual or single-setting.[52,72,73,102] **Developing**.
B. Formal Curricula		

Sub-domain	Score (codes)	Supporting evidence & synthesis
AI Literacy	**2 (L)**	The most consistently delivered GME content, with implemented residency examples (AI-RADS; DATA-MD) and radiology-resident curricula, but single-site; AI was not embedded in formal pediatric residency curricula despite growing clinical use.[18,19,68,70] **Developing**.
Clinical Application (incl. clinical decision support)	**2 (L)**	Widely described but unevenly delivered; the most concrete deployment is an ambient AI-scribe pilot (~1,000 notes across 31 internal medicine residents), evaluated qualitatively only; radiology curricula often over-weight technical model development relative to clinical use.[19,72,74] **Developing**.
Critical Appraisal of AI Tools	**2 (L)**	The best-evidenced GME curricular element: DATA-MD's hands-on appraisal workshop produced its largest knowledge gain (2.0→3.0, P<.001), AI-RADS centered a journal club on appraisal, and residents named appraisal their top need.[18,19,78] Upper **Developing**.
Specialty-Specific	**2 (L,G,P)**	Well documented but heavily concentrated in radiology (implemented curricula, society activity, board content); isolated examples in pediatrics, psychiatry, family medicine, and endoscopy, with a recurrent generalist gap.[18,68,72,76] **Developing**.
C. Faculty Development		
Qualified Faculty Availability	**1 (L)**	A barrier rather than a delivered program: an inversion of expertise (trainees outpacing supervisors) and low attending role-modeling (~10–22%); no evaluated GME faculty development program is documented.[32,67] **Emerging**.
Interdisciplinary Collaboration	**2 (L)**	Repeatedly recommended and partly operationalized: the DATA-MD development panel spanned clinicians, computer scientists, engineers, informaticians, and lawyers; an interprofessional ethics workshop spanned six fields, documented but not formalized into reproducible structures.[19,60,68] **Developing**.
Budget / Resources Allocated	**1 (L,G)**	Appears mainly as a barrier or incidental note (DATA-MD grant-funded; an institutional license purchased for the scribe pilot); no budgeted, systematic GME resourcing model.[19,68,74] **Emerging**.
D. Assessment		
Validated Methods / Instruments	**1 (L)**	Largely absent; reliance on locally developed, non-validated knowledge tests; a validated confidence instrument appears within an interprofessional workshop rather than a GME curriculum.[19,55,60] **Emerging.**
Milestones or EPAs	**1 (L,P)**	Not in force; conceptual work includes a preliminary 5-level ACGME-milestone-style framework, ACGME-domain mappings, and an EPA-for-AI concept within AI-PACE.[52,53,72,74] **Emerging**.
Summative vs. Formative	**1 (L)**	Predominantly formative and self-reported (pre/post knowledge checks; a formative self-checklist); no validated summative AI-competency assessment tied to high-stakes decisions.[18,19,30] **Emerging**.
Outcomes (e.g., patient) Data	**2 (L)**	Outcome data exist but are shallow: DATA-MD effect sizes 0.74–1.80; AI-RADS content-mastery gains and 9.8/10 satisfaction; positive effects in four of five radiology curricula; all are single-site, short-term, Level 1–2, with no behavior or patient outcomes.[18,19,60,68] **Developing**.
E. Ethics / Equity / Legal		
Policy & Governance	**2 (L,G)**	Broadly recognized content (a structured ethics workshop; dedicated radiology thematic areas; psychiatry regulatory mapping; a DATA-MD governance module that showed no

Sub-domain	Score (codes)	Supporting evidence & synthesis
		measurable gain), but unevenly implemented and weakly assessed.[19,60,72,75] **Developing**.
Algorithmic Bias	**2 (L)**	Explicitly and recurrently taught (DATA-MD's bias module citing the Obermeyer study; VALIDATE's bias step flagging pediatric subpopulations; a bias-and-fairness domain), but rarely tied to demonstrated competence.[19,30,60] **Developing.**
F. Accreditation, Licensing & Professional Guidance		
Accreditation Standards Requiring AI	**1 (P,L)**	No binding standard; sources repeatedly note AI's absence from ACGME requirements and recommend incorporating AI competencies into milestones as a future step (advocacy/anticipatory).[55,72,78,85] **Emerging**.
Licensing / Board Exam AI Content	**1 (P)**	Essentially limited to one concrete radiology example: the ABR added AI to the noninterpretive-skills section of its Qualifying (Core) Exam in 2021, against broad absence elsewhere (e.g., the USMLE).[68,85] **Emerging**.
Professional Society Guidance	**2 (G,P,L)**	Accumulating: an ESGE endoscopy curriculum, ACR/Royal College activity, and regulatory/WHO mapping, but largely advisory, uneven, and non-mandatory at the GME level.[68,72,75,76] **Developing**.
CC / AI Certification Pathway	**1 (L,G)**	No formal CC requirement or dedicated pathway; adjacent signals include a referenced AI-in-medicine certification, AI fellowships, and survey support for AI in CME; isolated and more CME- than GME-anchored.[47,72,83] **Emerging**.

Detailed 18-sub-domain matrix — CME

Sub-domain	Score (codes)	Supporting evidence & synthesis
A. Defined Competency Frameworks		
Defined Competency Frameworks	**2 (L,G)**	Several CME-relevant frameworks are proposed (ESGE statements; primary care domains; Russell's domains/subcompetencies; AI-PACE; a Miller's-Pyramid progression) but diverge and are uncoordinated; only two curriculum frameworks were identified in the reviewed literature, none theory-grounded.[47,52,55,76] **Developing**.
B. Formal Curricula		
AI Literacy	**2 (L,G)**	The best-covered curricular area, with named real CME offerings and fundamentals taught in ~79% of reviewed programs, but inconsistent depth and only one delivered study with measured outcomes (D'Souza: $4\pm1.4\rightarrow9\pm1.3/10$, $P<.0001$).[55,60,84,85] **Developing**.
Clinical Application (incl. clinical decision support)	**2 (L,G)**	Widely endorsed but rarely delivered-and-evaluated; the most concrete treatment is the ESGE position statement specifying phased hands-on computer-aided detection (CADe) training for practicing endoscopists; others describe application conceptually.[52,76,83,103] **Developing**.
Critical Appraisal of AI Tools	**2 (L)**	A strong recurring theme organized around an EBM analogy and ML users'-guide approach, but addressed in only ~37% of reviewed programs with no delivered, assessed appraisal curriculum.[55,63,103,104] **Developing**.
Specialty-Specific	**2 (L,G)**	Multiple specialty-anchored treatments at or near CME level (endoscopy, hematology/oncology, psychiatry, radiology attitudes, radiomics, primary care), but mostly surveys, reviews, or

Sub-domain	Score (codes)	Supporting evidence & synthesis
		proposals; ESGE is the one society-backed specialty CME curriculum.[47,75,76,82,83] **Developing**.
C. Faculty Development		
Qualified Faculty Availability	**1 (L)**	The most consistently named barrier and the least solved (only 14% of imaging lecturers had any formal AI training; current faculty lack an ML foundation); no implemented faculty development program is documented.[47,48,63,85] **Emerging**.
Interdisciplinary Collaboration	**2 (L,G)**	A prominent recommendation with one executed interprofessional design (six fields) and empirical workforce preference for clinician-plus-IT activities, but no outcome data linking collaboration structures to learner results.[52,60,83] **Developing**.
Budget / Resources Allocated	**1 (L)**	Mostly acknowledged as a barrier (faculty, infrastructure, licensing costs); one analysis offers a substantive cost-mitigation pathway (CPU-only open-source stacks, capped cloud/GPU, reusable case bundles); no documented resourced program with outcomes.[53,75,84,85] **Emerging**.
D. Assessment		
Validated Methods / Instruments	**1 (L)**	No widely adopted validated CME AI-competency instrument; MAIRS-MS recurs but its validity for measuring change is contested, and surveys measured perceptions, not skills.[48,55,83] **Emerging**.
Milestones or EPAs	**1 (L,P)**	Conceptual and structurally more native to UME/GME; AI-PACE uniquely engages EPAs/entrustment for AI and Russell maps to ACGME domains, but the CME analog (CC/CME-credit competency assessment) is essentially absent.[47,52,53] **Emerging**.
Summative vs. Formative	**1 (L)**	Touched only briefly; one workshop's pre/post MCQ functions summatively, and formative/adaptive possibilities are noted conceptually, with no systematic architecture for CME.[60,92,102] **Emerging**.
Outcomes (e.g., patient) Data	**1 (L)**	Learning outcomes exist for one workshop only; reviewed programs reached Kirkpatrick Levels 1–2 with none at 3–4, and behavior/patient-outcome data are absent.[48,53,55,60] **Emerging**.
E. Ethics / Equity / Legal		
Policy & Governance	**2 (L,P,G)**	Well represented via external policy positions (AMA augmented-intelligence; ACP; NAM AI Code of Conduct; WHO/EU AI Act/FDA SaMD) and governance proposals, but referenced secondhand rather than embedded as standardized CME; one ethics workshop delivered measured gains.[60,75,84,85] **Developing**.
Algorithmic Bias	**2 (L)**	One of the most consistently present content areas across nearly the whole corpus (bias subcompetencies; skin-type fairness; debiasing; named fairness tools; cognitive biases), but with no implemented debiasing curriculum with outcomes.[47,52,75,76,84,104] **Developing**.
F. Accreditation, Licensing & Professional Guidance		
Accreditation Standards Requiring AI	**1 (L,P)**	Addressed primarily to document absence: neither LCME nor ACGME holds binding AI standards (per a 2025 Josiah Macy/AAMC/ACGME recommendation), and no AI accreditation requirements exist; one emerging non-physician standard requires AI awareness for UK radiographers.[48,53,55,85] **Emerging**.
Licensing / Board Exam AI Content	**1 (L,P)**	Mainly documented absence and proposal: the USMLE does not test AI/data-science content; LLMs taking exams are a category mismatch, not physician AI-competency content.[84,85,92] **Emerging**.

Sub-domain	Score (codes)	Supporting evidence & synthesis
Professional Society Guidance	**2 (G,L,P)**	The most mature accreditation-adjacent domain: a formal society CME curriculum (ESGE) plus guidance from Canadian/European/UK imaging bodies, Josiah Macy/AAMC/ACGME, AAFP/CFPC/ABF/AMIA, AMA, ACP/NAM, ASCO/ASH, and RSNA CME, reflecting extensive society engagement though advisory and uneven, consistent with Developing.[47,48,53,76,83,84] **Developing**.
CC / AI Certification Pathway	**1 (L,G)**	Only adjacent standalone offerings (an AI-in-medicine certification; Harvard/Stanford/AMIA programs; a CME course catalogue), with no continuing certification AI requirement tied to board recertification.[47,84,85] **Emerging**.

Appendix D. Evidence Traceability — Consolidated Scores to Reference Numbers

Each consolidated six-domain score maps to the reference numbers that justify it, so every score is traceable. The numbers below are the union of the references cited in the basis column of that reporting domain across the UME, GME, and CME consolidated matrices (Sections 3.3, 4.4, and 5.4), and key to the master reference list (Appendix A).

Reporting domain	UME	GME	CME	Justifying references (by reference number)
1. Defined competency frameworks	**2**	**2**	**2**	47,51,52,53,55,72,73,76
2. Formal curricula	**2**	**2**	**2**	18,19,35,36,54,55,56,60,68,74,84,85
3. Faculty development	**1**	**1**	**1**	22,32,34,47,48,57,58,59,60
4. Validated assessment methods	**1**	**1**	**1**	18,19,51,53,55,59,60,61,72
5. Ethics / equity / legal	**2**	**2**	**2**	19,46,52,60,62,75,76,77,84
6. Accreditation / licensing / board exam / society guidance / CC	**1**	**1**	**1**	46,47,48,55,63,64,68,76,78,84

Appendix E. Glossary of Key Terms, Abbreviations, and Organizations

This glossary defines common medical-education terms, abbreviations, and organizations used in this report, together with widely used educational and evaluation frameworks included as background for readers.

Stages of medical education and certification

UME — Undergraduate Medical Education; the medical-school phase leading to the MD or DO degree.

GME — Graduate Medical Education; post-medical-school clinical training in residency and fellowship programs.

CME — Continuing Medical Education; ongoing education that helps practicing physicians maintain and update their knowledge and skills.

CC — Continuing Certification; the ongoing requirements set by specialty certifying boards for physicians to keep their board certification current.

Care-delivery and education paradigms

LHS — Learning Health System; a health system, and the infrastructure, data, and culture supporting it, in which routine clinical data are continuously analyzed to generate knowledge that is fed back to improve care, increasingly with AI.

PM — Precision Medicine; an approach to care that tailors prevention, diagnosis, and treatment to the individual patient using data, algorithms, and molecular and computational tools.

PME — Precision Medical Education; application of the learning-health-system data-to-practice cycle to physician training, using data and AI to personalize teaching, assessment, and coaching to the individual learner (also termed precision education).

Competency, curriculum, and assessment terms

Competency — an observable ability of a physician that integrates knowledge, skills, values, and attitudes; competencies develop over time, can be assessed, and are the building blocks of a competency framework.

Competency framework — a structured set of competencies, and the behaviors that demonstrate them, defining what a learner should be able to do; the underlying approach is often called competency-based medical education (CBME).

Milestones — observable, developmentally sequenced markers of progress within a competency; used notably by the ACGME to describe and track resident development over time.

Entrustable professional activity (EPA) — a defined unit of professional practice that can be entrusted to a trainee once they demonstrate enough competence to perform it at a specified level of supervision.

Core EPAs — the AAMC's set of activities that graduating medical students are expected to be ready to perform on entering residency; the report discusses proposed AI-related additions to this set.

Formative assessment — assessment "for learning": lower-stakes evaluation that provides feedback to guide improvement.

Summative assessment — assessment "of learning": higher-stakes evaluation that judges achievement against a standard.

Required vs. elective curriculum — content that all learners must complete, versus optional content they may choose.

Longitudinal (spiral) vs. bolt-on curriculum — content revisited and progressively deepened across a program, versus a single stand-alone module added to an existing curriculum.

Clerkship — a supervised clinical rotation in medical school in which students take part in patient care within a specialty.

CanMEDS — the competency framework of the Royal College of Physicians and Surgeons of Canada, organized around physician roles; CanMEDS-FM is its family-medicine adaptation.

RACI matrix — a responsibility-assignment chart labeling each party as Responsible, Accountable, Consulted, or Informed; used in the report's GME stakeholder mapping.

Educational and evaluation frameworks

Bloom's taxonomy — a hierarchical classification of learning objectives used to write and sequence educational goals; in its widely used revised cognitive form it ascends from remembering and understanding, through applying and analyzing, to evaluating and creating, and it helps educators pitch objectives at the intended level of thinking.

Miller's pyramid — a four-level framework for assessing clinical competence that progresses from knows (recall of facts) to knows how (application of knowledge) to shows how (demonstration in a controlled setting such as a simulation or OSCE) to does (performance in independent practice).

Master Adaptive Learner (MAL) — a model of self-regulated, metacognitive learning in which the clinician cycles through planning, learning, assessing, and adjusting in order to develop adaptive expertise: the capacity to perform well in novel or unfamiliar situations rather than only in routine ones.

Kirkpatrick levels — a four-level scheme for evaluating educational programs, ascending from Level 1 (reaction, such as learner satisfaction) and Level 2 (learning, such as knowledge, skills, or attitudes) to Level 3 (behavior change in practice) and Level 4 (results, including patient and system outcomes); this report uses these levels to describe the ceiling of the current evidence.

Accrediting, certifying, licensing, and professional organizations

LCME — Liaison Committee on Medical Education; accredits MD-degree-granting medical-education programs in the United States and Canada.

ACGME — Accreditation Council for Graduate Medical Education; accredits US residency and fellowship programs and maintains the Milestones framework.

ACCME — Accreditation Council for Continuing Medical Education; accredits organizations that provide CME to physicians.

AAMC — Association of American Medical Colleges; represents MD-granting medical schools and teaching hospitals; co-administers the Curriculum SCOPE Survey and publishes the Core EPAs and the New and Emerging Areas competency series.

AACOM — American Association of Colleges of Osteopathic Medicine; represents DO-granting (osteopathic) medical colleges; co-administers the Curriculum SCOPE Survey.

ABR — American Board of Radiology; the certifying board for radiology, cited for incorporating AI into its examinations.

FSMB — Federation of State Medical Boards; the national organization of the state boards that license physicians, and co-sponsor of the USMLE with the NBME.

NBME — National Board of Medical Examiners; an independent, not-for-profit organization that develops and administers assessments for medical students, residents, and physicians, and that co-sponsors the United States Medical Licensing Examination (USMLE) with the FSMB.

AMA — American Medical Association; the largest US physician professional association, author of AI policy and, with DiMe, the clinical practice landscape report used here.

NAM — National Academy of Medicine; an independent health-and-medicine advisory body within the National Academies (formerly the Institute of Medicine) and convener of the Clinicians Forum on AI cited here.

National Academies (NASEM) — the National Academies of Sciences, Engineering, and Medicine, the umbrella body (which includes the NAM) that published the 2023 workshop proceedings on AI in health professions education.

WHO — World Health Organization; the United Nations public-health agency, cited for AI-in-health guidance.

ESGE — European Society of Gastrointestinal Endoscopy; a specialty society cited for a formal society-level CME curriculum on AI.

AMIA — American Medical Informatics Association; the US professional association for biomedical and health informatics.

ACP — American College of Physicians; the professional society of internal medicine.

AMEE — Association for Medical Education in Europe; an international medical-education association and a collaborator on the IACAI framework.

AAHCI — Association of Academic Health Centers International; a global organization of academic health centers and a collaborator on the IACAI framework.

IAMSE — International Association of Medical Science Educators; an international society for medical-science teaching and a collaborator on the IACAI framework.

IACAI — International Advisory Committee for Artificial Intelligence; source of the AI Integration Framework (Learner and Educator matrices and GME stakeholder/RACI tables) cited in the report as AAMC/MedBiquitous instruments.

MedBiquitous — an AAMC program that develops technology standards for health-professions education; associated with the IACAI framework.

DiMe — Digital Medicine Society; a nonprofit professional society for digital medicine and co-author of the AMA–DiMe clinical practice landscape report.

ABMS — American Board of Medical Specialties; the umbrella organization that oversees the 24 US member specialty certifying boards and Continuing Certification (CC) standards.

Josiah Macy Jr. Foundation — a US foundation dedicated to improving the education of health professionals; funder of the Macy Foundation Innovation Reports Part I and Part II on AI in medical education cited in this analysis.

Artificial-intelligence and data terms

AI — Artificial Intelligence; computer systems that perform tasks ordinarily requiring human intelligence.

ML — Machine Learning; a branch of AI in which systems learn patterns from data rather than following explicitly programmed rules.

Generative AI — a class of AI systems, typically built on large machine-learning models, that produce new content (text, images, audio, or code) in response to prompts rather than only classifying or predicting from existing data; LLMs are the generative-AI systems most used in medical education and clinical care.

LLM — Large Language Model; an AI model trained on large text corpora to interpret and generate natural language, and the basis of many generative-AI and consumer chatbot tools.

EHR — Electronic Health Record; the digital record of a patient's health information used in clinical care.

EBM — Evidence-Based Medicine; integrating the best available research evidence with clinical expertise and patient values.

Predictive analytics / risk models — algorithms trained on clinical data to estimate the likelihood of a future event (e.g., hospital readmission, sepsis, deterioration) and support prioritization and triage.

Imaging AI and pattern-recognition tools — AI systems, typically deep-learning based, that analyze medical images (radiology, pathology, ophthalmology, dermatology) to detect, classify, or measure clinically relevant findings.

Clinical decision support (CDS) — software that presents evidence-based recommendations, alerts, or contextual information to clinicians at the point of care to inform diagnosis, treatment, or monitoring decisions; may or may not incorporate machine learning.

Ambient documentation / AI scribes — AI tools that passively capture the audio of a clinical encounter and generate a draft clinical note, order set, or after-visit summary for physician review and editing.

Patient-facing AI tools — AI systems used directly by patients (e.g., symptom checkers, chatbots for triage or coaching, patient-education generators) that may influence care-seeking or self-management before or between physician encounters.

Agentic or workflow-executing AI systems — AI systems that go beyond producing recommendations to autonomously execute multi-step tasks in clinical workflows (e.g., scheduling, ordering, drafting communications, coordinating referrals), typically with defined human-in-the-loop checkpoints.

Appendix F. Possible Clinical Use Cases of AI in Practice

The categories below map to the AI-terms glossary in Appendix E. The specific tools named are illustrative rather than exhaustive, and inclusion here is not an endorsement; regulatory status, evidence base, and appropriate use vary by product, setting, and patient population.

Predictive analytics and risk models

- **Deterioration and early-warning scores.** Continuous or near-continuous scoring of hospitalized patients for sepsis, clinical deterioration, or readmission risk, generating alerts that support triage and escalation.
- **Length-of-stay and discharge planning.** Estimating expected length of stay or discharge disposition to support care coordination, transitions, and resource allocation.
- **Chronic-disease risk stratification.** Prioritizing panels of primary care patients (e.g., cardiovascular, diabetes, kidney disease) for outreach, preventive care, or specialty referral.

Imaging AI and pattern-recognition tools

- **Radiology triage and prioritization.** Detection of findings such as intracranial hemorrhage, pulmonary embolism, or large-vessel occlusion on imaging to reorder worklists and shorten time to treatment.
- **Screening and interpretation support.** Assistance with mammographic, chest-radiographic, and diabetic retinopathy screening, and quantitative measurements in cardiac and musculoskeletal imaging.
- **Pathology and dermatology.** Whole-slide-image analysis for cancer detection and grading; skin-lesion classification to support referral decisions.

Clinical decision support (CDS)

- **Diagnostic and treatment recommendations.** Evidence-linked suggestions at the point of care (e.g., antibiotic selection, anticoagulation, chemotherapy regimen selection) integrated into electronic health record (EHR) workflows.
- **Medication safety.** Real-time checks for drug–drug interactions, contraindications, dose adjustments, and duplicate therapy at ordering.
- **Guideline-concordance and quality alerts.** Prompts to close care gaps (e.g., cancer screening due, immunization due) and to align care with current guidelines.

Generative AI and large language models (LLMs)

- **Point-of-care clinical decision support.** Case-specific questions posed to a general-purpose or clinically tuned large language model (LLM) to generate or broaden a differential diagnosis, stress-test a working diagnosis, surface guideline-based management options, or serve as a curbside-consultation substitute, with the physician responsible for verifying outputs against primary sources before acting on them.
- **Clinical documentation drafting.** Draft progress notes, discharge summaries, and referral letters generated from structured data and prior notes for physician review and editing.
- **Patient communication drafting.** Draft responses to patient portal messages, patient-education explanations, and after-visit summaries tailored to reading level, revised and signed by the physician.
- **Knowledge retrieval and synthesis.** Rapid literature synthesis, guideline lookup, and case-specific summarization to support clinical reasoning and education.
- **Coding and billing assistance.** Suggested diagnosis and procedure codes derived from clinical documentation for physician confirmation.

Ambient documentation and AI scribes

- **Ambient encounter capture.** Passive capture of the clinician–patient conversation to generate a structured draft note, orders, and follow-up tasks for physician review at or after the visit.
- **Multi-specialty adaptation.** Specialty-tuned scribe outputs (e.g., procedural notes, inpatient rounds, primary care visits) that respect specialty documentation conventions.

Patient-facing AI tools

- **Symptom triage and self-assessment.** Chatbot-style tools that help patients decide when and where to seek care, and prepare information for the encounter.
- **Chronic-disease self-management.** Conversational coaching and monitoring for conditions such as diabetes, hypertension, and mental-health symptoms, with pathways to escalate to the care team.
- **Patient education.** AI-generated, plain-language explanations of diagnoses, medications, procedures, and results, adapted to the patient's stated reading level and language.

Agentic or workflow-executing AI systems

- **Multi-step workflow execution.** Systems that draft, route, and follow up on referrals, prior authorizations, laboratory orders, or scheduling with defined human-in-the-loop approval points.
- **Care-coordination agents.** Agents that reconcile medication lists across sites of care, arrange post-discharge follow-up, or coordinate cross-specialty consultations with physician oversight.
- **Population-health outreach.** Agents that identify patients due for preventive care, draft outreach communications, and schedule appointments, with clinician review before contact for higher-risk populations.

Across categories, the physician's accountable role remains: appraising fit for the case in front of them, disclosing AI involvement to patients as appropriate (i.e., when it materially affected the recommendation or record), and ensuring that AI outputs are reviewed rather than accepted by default. The educational implication is that physician AI readiness must be developed across all seven categories rather than being scoped narrowly to imaging or generative-AI tools.

Appendix G. Key Stakeholders for Physician AI Readiness

This analysis is relevant to, and its recommendations will require coordinated action from, the following stakeholders, each of whom has an essential role in ensuring that physicians are adequately prepared to use AI in clinical practice:

- Health professions educators and faculty across undergraduate medical education (UME), graduate medical education (GME), and continuing medical education (CME), including course, clerkship, and program directors
- Medical students, resident and fellow physicians, and practicing physicians
- Interprofessional colleagues, including nursing, pharmacy, and allied health professionals, whose roles and accountability shift alongside physicians' as AI is integrated into team-based care
- University, medical school, and health system leadership, including deans, designated institutional officials, and executives who set strategy, allocate resources, and govern AI deployment
- Accrediting, licensing, and certifying bodies, including the Liaison Committee on Medical Education (LCME), the Accreditation Council for Graduate Medical Education (ACGME), the Accreditation Council for Continuing Medical Education (ACCME), the National Board of Medical Examiners (NBME), the Federation of State Medical Boards (FSMB) and state licensing authorities, and specialty certifying boards
- Professional societies and specialty organizations that issue guidance and deliver education
- Policy makers and regulators at the federal and state levels
- Computer scientists, engineers, informaticists, and learning scientists who design, validate, and study both the tools and the curricula
- Industry, including AI developers, electronic health record vendors, and other health-technology companies
- Medical librarians and health-sciences informationists, who curate authoritative resources, support critical appraisal, and increasingly help learners evaluate AI-generated information
- Funders and philanthropies that finance curriculum development, faculty development, and outcomes research
- Patients, families, and patient communities, whose trust, safety, and outcomes are the end goal of physician AI readiness

Appendix H. Literature Search and Analytic Strategy

This appendix documents the literature search strategy, information sources, eligibility criteria, and the process used to screen, analyze, and consolidate the evidence base for this landscape analysis. The strategy used to collect primary stakeholder data through interviews and structured surveys is documented separately in Appendix I.

Search Strategy and Information Sources

Two databases—PubMed and Science.gov—were searched from inception to May 18, 2026, to identify articles addressing the medical education landscape for preparing physicians and trainees to use AI effectively in clinical practice. The PubMed search combined controlled-vocabulary MeSH terms with title keywords; no date or other limits were applied. A set of sentinel articles identified in advance was used to develop and test the search terms, and the strategy—developed first in PubMed—was translated as appropriate for Science.gov. Records were deduplicated in Covidence. The complete search strategies for both databases are provided below.

PubMed (searched May 18, 2026)—889 records. Search strategy:

(((("Artificial Intelligence"[MeSH Major Topic] OR "Artificial Intelligence"[ti] OR "AI"[ti] OR "Machine Learning"[ti] OR "Generative AI"[ti] OR "Large Language Models"[ti] OR "ChatGPT"[ti])) AND (("Medical Education"[ti] OR "Health Professions Education"[ti] OR "Undergraduate Medical Education"[ti] OR "UME"[ti] OR "Graduate Medical Education"[ti] OR curricul*[ti] OR "GME"[ti] OR "CME"[ti] OR "Faculty Development"[ti] OR "Medical Training"[ti] OR "Physician Education"[ti] OR "Resident Training"[ti]))))

Science.gov (searched May 18, 2026)—1,097 records. Search strategy: AI AND "Medical Education".

Across both databases, the searches returned 1,986 records in total. After 47 duplicates were removed in Covidence, 1,939 unique citations proceeded to screening.

Eligibility Criteria

Population. *Include:* medical students, residents, fellows, interns, physicians, and medical faculty across undergraduate, graduate, and continuing medical education; clinician-educators, faculty developers, and health care leadership; and interprofessional health care teams when physicians or medical students were part of the learning cohort. *Exclude:* patients, the general public, and caregivers; veterinary students; and non-clinical technical students such as pure computer-science, bioinformatics, or data-science majors, unless participating in an explicit interprofessional health care team training initiative.

Intervention. *Include:* AI attitudes, education, training, curricula, and resources; machine learning, large language models, machine intelligence, generative AI, and chatbots; clinical decision support; interprofessional education; training in clinical and diagnostic reasoning; prompt-engineering training; and teaching the critical appraisal of AI performance. *Exclude:* automated grading, AI tutors, test-item generation, feedback systems, and administrative efficiency in medical education; virtual patients used for general communication training, where AI is the patient rather than the clinical tool; predictive models for student or learner attrition; and studies focused solely on model performance, such as accuracy, sensitivity, and specificity.

Comparator. *Exclude:* non-medical professional settings and general workplace automation.

Outcome. *Include:* AI-related knowledge, skills, and digital-health literacy; attitudes, competency, AI literacy and proficiency, and readiness; preparation, experience, navigation, evaluation, teaching, courses, learning, and curricula; faculty development, clinical competence, and competency-based education; AI and algorithm aversion, AI trust, and trust calibration; prompt engineering and critical appraisal of AI; and entrustable professional activities, milestones, and workflow integration and adaptation. *Exclude:* technical model-performance metrics, institutional financial outcomes, and general-public perceptions of AI in society.

Screening and Selection

The 1,939 unique citations were imported into Covidence, where two reviewers independently screened every title and abstract. Seven hundred forty-eight articles advanced to a second round of title-and-abstract review, with disagreements resolved by a third reviewer. Sixty-six articles met the inclusion criteria at full-text review and were retained for analysis.

Evidence Analysis and Consolidation

The 66 included articles were organized in a shared Zotero library and sorted by education level into undergraduate, graduate, and continuing medical education folders; articles addressing more than one level, such as both undergraduate and graduate medical education, were filed under each. Within each level, articles were grouped into batches of three to five (the graduate medical education level, for instance, comprised nine batches).

Claude Opus was used to generate an analytic prompt, which was then applied—at the model's high, extra, and max settings—to analyze the batches at each level. Claude was also prompted to produce a second prompt for consolidated, level-specific (e.g., UME) analyses and syntheses; the batch analyses for a given level were uploaded to Claude and combined with this prompt to yield landscape analyses at the UME, GME, and CME levels.

An additional prompt was then generated, using Claude, to synthesize the evidence across all three levels. The three level-specific analyses were uploaded to Claude and combined into an initial consolidated landscape analysis.

This initial analysis was critically reviewed by hand for accuracy, depth, and relevance. Additional evidence and resources—including web-based courses, conference proceedings, other landscape analyses, and further relevant literature—were incorporated, and in some cases prompted revisions to the domain scores for a given level.

The analysis then underwent iterative, detailed non-systematic manual review and verification by the core team, who verified every citation and claim (including checking against source documents) and added or removed material on the basis of their subject-matter expertise in AI in health professions education and in a manner consistent with the aims of the analysis. The core team met weekly to every other week to discuss the report and resolve any discrepancies.

Claude Opus and Claude Fable were also prompted iteratively to assist with editing, including adding information and resources and performing spelling, grammar, and accuracy checks.

All prompts, batch analyses, and level-specific landscape analyses are available upon reasonable request.

Appendix I. Stakeholder Interviews and Surveys

This appendix documents the strategy used to elicit qualitative stakeholder feedback from leaders across medicine and medical education, together with the structured surveys of graduate medical education program directors and continuing professional development directors whose findings are reported throughout this analysis.

Interview Strategy

Dr. Burstin reached out to leaders of national organizations closely affiliated with medical education across the continuum to assess their perspectives on the current state of AI readiness for physicians in undergraduate medical education (UME), graduate medical education (GME), and continuing medical education (CME), and on potential next steps to upskill the clinical workforce. In-person and virtual interviews were conducted across education organizations spanning the continuum, including the Association of American Medical Colleges (AAMC), the Accreditation Council for Graduate Medical Education (ACGME), the Accreditation Council for Continuing Medical Education (ACCME), the American Medical Association (AMA), the Coalition for Health AI (CHAI), the Digital Medicine Society (DiMe), and other national leaders of related organizations. Interviews were transcribed and summarized by Dr. Burstin. Salient points from the interviews were extracted for inclusion in the broader report.

Survey Strategy

To better assess the perspectives of frontline educators in GME and CME, the Council of Medical Specialty Societies (CMSS) solicited structured surveys from two sources.

Organization of Program Director Associations (OPDA). OPDA is convened by CMSS to promote the role of residency and fellowship program directors and societies in achieving excellence in graduate medical education. OPDA includes 30 program director associations across specialties (https://cmss.org/programs-and-resources/opda/).

Program director input was solicited to better understand how residency programs are preparing residents for the use of artificial intelligence (AI) in clinical practice, including both generative AI and clinically deployed predictive AI tools.

Program director association representatives were asked to share the survey with their program directors using either of two approaches: surveys were sent to residency program directors with a particular interest in the effective use of AI during residency, or surveys were distributed broadly to all program directors in their specialty.

The OPDA survey instrument is reproduced in Appendix J.

CMSS Continuing Professional Development Directors. CMSS convenes a Professional Peer Group (PPG) of Continuing Professional Development (CPD) Directors across CMSS's 58 member societies. The CPD PPG provides an opportunity to network, exchange ideas, and share concerns on key issues related to CPD, and serves as a recognized forum for CME and continuing education (CE) directors to voice their positions and concerns on issues affecting the delivery and conduct of CME and CE.

The CPD-Focused CMSS Physician AI Readiness Survey was sent to all society education directors to understand how medical specialty societies are addressing artificial intelligence (AI) in their continuing medical education (CME) programming.

The CPD survey instrument is reproduced in Appendix K.

Respondents and Reporting Conventions

The OPDA program director survey was completed by 114 respondents. The CPD-Focused CMSS Physician AI Readiness Survey was completed by 34 respondents representing 28 different member societies. Findings from both surveys, together with salient points from the stakeholder interviews, are integrated into the level-specific and cross-cutting sections of this report and are identified there as primary data collected for this analysis. Percentages reported in the text are calculated on the total number of respondents to the relevant survey; several items permitted respondents to select more than one option, so percentages for those items do not sum to 100. Consistent with the evidence conventions in Appendix B, these primary data are treated as source code O and corroborate, but do not change, the consolidated domain scores.

Appendix J. OPDA Program Director Survey Instrument

The survey instrument distributed to residency and fellowship program directors through the Organization of Program Director Associations (OPDA) is reproduced below. The strategy governing its development and distribution is described in Appendix I, and its findings are reported in Sections 2, 4.3, 4.4, 7.1, 8.1, and 8.4.

1. Are your residents currently using AI tools in their clinical work or clinical learning? Select all that apply. *(required; multi-select)*

a. No
b. Yes, for clinical documentation (e.g., ambient scribes) or administrative tasks (e.g., chart summarization)
c. Yes, for clinical decision support using commercial LLMs
d. Yes, for clinical decision support using models embedded in AI
e. Yes, for patient communication or education
f. Yes, for learning or board preparation
g. Unsure
h. Other (open text)

2. What are the most significant barriers to preparing residents to use AI safely and effectively in clinical practice? Select up to three *(required; multi-select, max 3)*

a. Limited faculty expertise
b. Limited curriculum time
c. Limited access to relevant AI tools
d. Lack of standardized competencies
e. Lack of assessment methods
f. Limited institutional guidance or governance
g. Ethical, legal, or privacy concerns
h. Limited evidence about effective educational approaches
i. Competing educational priorities
j. Unsure
k. Other (open text)

3. What training is most critical for residents to be prepared for AI-enabled clinical practice? *(required; multi-select)*

a. AI literacy or basic AI concepts
b. Clinical applications of AI
c. Critical appraisal of AI tools
d. Recognizing AI limitations, errors, or bias
e. Ethics, equity, legal, or privacy issues
f. Human-AI collaboration in clinical care
g. AI has not been formally addressed
h. Unsure

4. How confident are you that your residents can effectively appraise AI outputs for accuracy, bias, safety, and contextual appropriateness for their patients? *(required; single-select)*

a. Not confident
b. Slightly confident
c. Moderately confident
d. Very confident
e. Unsure

5. What training would most improve your residents' ability to use AI tools safely and critically? *(required; open text)*

6. What type of AI knowledge, skills, and attitudes should your entering and/or graduating residents have? *(required; open text)*

Respondent information

a. Name (optional): First Name / Last Name

b. Email (optional)
c. Specialty (optional)

Appendix K. CPD-Focused CMSS Physician AI Readiness Survey Instrument

The continuing professional development (CPD)-focused Council of Medical Specialty Societies (CMSS) Physician AI Readiness Survey instrument, distributed to the education directors of CMSS member societies, is reproduced below. The strategy governing its development and distribution is described in Appendix I, and its findings are reported in Sections 2, 5.3, 5.4, 5.5, 6, 7.1, 8.1, 8.2, and 8.4.

1. Does your society offer continuing medical education (CME) programs for your members that address the safe and effective use of AI in clinical practice? *(required; single-select)*

a. No active AI-focused programs
b. Some AI content is incorporated into existing programs
c. AI-focused programs are under development
d. AI-focused programs for CME and/or MOC credits
e. Unsure

2. How would you characterize your level of confidence in your society's ability to educate physicians to safely and responsibly use AI in practice? *(required; single-select)*

a. Not confident
b. Slightly confident
c. Moderately confident
d. Very confident
e. Unsure

3. What are the most significant barriers and resource constraints that limit your ability to develop CME programs that could better prepare physicians to use AI safely and critically in clinical practice? Select up to three. *(required; multi-select, max 3)*

a. Limited faculty expertise
b. Lack of standardized competencies
c. Lack of assessment methods
d. Ethical, legal, or IP concerns
e. Limited evidence about effective educational approaches
f. Limited ability to assess learner or physician AI competencies
g. Competing educational priorities
h. Unsure
i. Other (open text)

4. What content would be most necessary to ensure members of your society are prepared for AI-enabled clinical practice? *(required; multi-select)*

a. AI literacy or basic AI concepts
b. Clinical applications of AI
c. Critical appraisal of AI tools
d. Recognizing AI limitations, errors, or bias
e. Ethics, equity, legal, or privacy issues
f. Human-AI collaboration in clinical care
g. Unsure

5. Have you gauged your members' interest in CME content that would prepare them to use AI safely and effectively in clinical practice? *(required; single-select)*

a. Yes
b. No

6. Please feel free to add any additional comments regarding AI-focused CME *(optional; open text)*

Respondent information

a. Name (optional): First Name / Last Name
b. Email (optional)

c. Society (required; dropdown of CMSS member societies)