

RESEARCH

Open Access



A research roadmap for AI opportunities in student assessment for medical education

Morteza Rezaei-Zadeh^{1*} and Magdalena Cerbin-Koczorowska²

Abstract

The integration of Artificial Intelligence (AI) in medical education is rapidly transforming assessment practices, offering unprecedented opportunities to enhance student evaluation, feedback, and learning pathways. However, despite the potential, a comprehensive understanding of these opportunities and their interdependencies has been lacking. This study provides a critical review of the literature on AI's role in medical education assessment, categorising 22 identified opportunities into seven major "mega-opportunities" that address various aspects of student assessment. Through the application of Interpretive Structural Modelling (ISM), the cause-effect interdependencies among these mega-opportunities were explored, revealing a complex web of relationships that guide their effective implementation. The findings highlight the central role of "Automated Feedback and Evaluation" and "Data-Driven Analytics and Curriculum Improvement" as foundational drivers, with far-reaching impacts on other areas like "Simulation-Based Assessment" and "Longitudinal Assessment and Development." This paper culminates in the proposal of a *research roadmap* that highlights the priority of addressing different mega-opportunities in AI and assessment, offering practical guidelines for medical researchers, educators, institutions, and policymakers to adopt AI-driven assessment strategies. Future research avenues are identified to explore the real-world application and impact of these AI-driven innovations, focusing on longitudinal studies and educational equity. The findings underscore the need for continued research to refine the model proposed by this study and adapt it to diverse educational environments.

Keywords Artificial intelligence, Medical education, Student assessment, AI opportunities

Introduction

Artificial intelligence (AI) is reshaping the landscape of student assessment, offering novel opportunities for enhancing accuracy, efficiency, and personalisation across various stages of the assessment process - impacting students, markers, moderators, and institutional bodies such as boards of studies. In higher education, AI is increasingly integrated into a range of assessment

practices. This includes summative assessments - such as automated scoring and adaptive testing [1]; formative approaches, including real-time feedback systems; and programmatic strategies, where machine learning-powered predictive analytics can identify students at risk of underperformance with up to 88% accuracy, enabling early intervention and personalised support [2]. In this article, we understand AI broadly as algorithmic computer systems capable of performing tasks that typically require human intelligence [3], applied across these three assessment domains. As the global AI in education market is projected to surpass from USD 4.8 billion in 2024 to USD 75 billion by 2033 [4], these innovations signal a profound shift in how educators approach data collection

*Correspondence:

Morteza Rezaei-Zadeh

mrz5@leicester.ac.uk

¹University of Leicester, Leicester, UK

²University of Edinburgh Medical School, Edinburgh, United Kingdom



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

and utilisation to support personalised growth and make reliable judgements about student's performance—raising both exciting possibilities and complex ethical considerations.

These issues received considerable critical attention within the field of medical education (ME), where decisions concerning students' competence and readiness for practice carry profound implications for addressing healthcare needs and patient safety. Despite significant progress observed in ME assessment practices over recent decades, training programs continue to encounter challenges that hinder effective assessment practices [5].

Traditional assessment methods - such as multiple-choice quizzes, written essays, and even clinical evaluations - often fall short in capturing the breadth and depth of competencies required for medical practice [6]. These methods tend to focus heavily on knowledge recall, frequently neglecting higher-order cognitive skills such as clinical reasoning, communication, and ethical decision-making [7]. Furthermore, issues like subjectivity in scoring, variability among assessors, and lack of timely, actionable feedback persist across many medical education contexts [8]. Given these longstanding challenges, various stakeholders emphasise the urgent need for more adaptive, scalable, and personalised approaches to assessment using AI [9]. While AI presents transformative opportunities, some caution against overreliance on EdTech in sensitive areas like education and assessment [10].

Recent advances in AI offer promising solutions to many of the above-mentioned challenges. AI-powered tools, including Natural Language Processing (NLP), Machine Learning (ML) algorithms, and Intelligent Tutoring Systems (ITS), are increasingly being used to provide real-time, individualised feedback, simulate clinical decision-making environments, and evaluate complex learner behaviours [11, 12]. Additionally, AI supports scalability by automating grading and offering adaptive testing models that respond to the learner's performance in real-time [13]. These capabilities are seen as a potential opportunity to improve the precision and fairness of assessments while aligning with the broader shift toward data-driven precision education [14]. Despite the growing body of research exploring AI in medical education, current literature on AI's role in student assessment remains fragmented and descriptive, often focusing on isolated use cases or theoretical potential without synthesising broader trends or outcomes [15, 16]. While several studies highlight specific opportunities - such as automated scoring, adaptive testing, or diagnostic simulations - few offer a comprehensive framework that connects these developments or evaluates their interdependencies. Moreover, existing reviews often overlook the strategic implications of AI for long-term assessment reform, especially in relation to competency-based education and ethical evaluation practices [17]. This gap calls for a systematic, critical review that

not only maps existing opportunities but also models their causal relationships and provides a future-oriented roadmap for research and practice.

Building upon the identified gap, this paper aims to provide a comprehensive and critical review of the literature on the opportunities that AI offers for student assessment in medical education. By systematically categorising existing studies and modelling the interdependencies among the reported opportunities, the study seeks to move beyond descriptive analysis and offer an integrative understanding of how AI is reshaping assessment practices in medical education. In doing so, it also offers a forward-looking roadmap to guide future research and practical implementation. To achieve these aims, this study addresses the following research questions (RQs):

RQ1: What opportunities does AI offer for student assessment in medical education?

RQ2: What are the interdependencies and structural relationships among these identified opportunities?

RQ3: How can these relationships be leveraged to develop a strategic roadmap for implementing AI in medical education assessment?

The paper is structured as follows: First, it outlines the methodological approach used in the literature review and modelling. Next, it presents the key categories of AI-driven assessment opportunities identified across the literature. This is followed by a discussion of the interrelationships among these categories and their broader implications. The paper concludes with a critical reflection on existing gaps and offers recommendations for future research directions and policy considerations.

Methodology

This study adopts a qualitative, exploratory research design [18], employing a Systematic Literature Review (SLR) complemented by thematic synthesis and causal modelling using Interpretive Structural Modelling (ISM) to explore and conceptualise the opportunities that AI brings to student assessment in medical education.

Research design

Given the dynamic and rapidly evolving nature of AI technologies in education, particularly in student assessment, an SLR was selected as the primary method to synthesise existing knowledge. To further explore the interrelationships among the identified opportunities, the synthesis was followed by the application of ISM. ISM is a well-established method used to analyse complex systems by identifying and structuring interdependencies among elements. Its relevance in educational and healthcare research lies in its ability to construct hierarchical models that help visualise how certain opportunities may act as enablers for others, providing strategic insights for implementation [19, 20]. In this study, ISM was used to

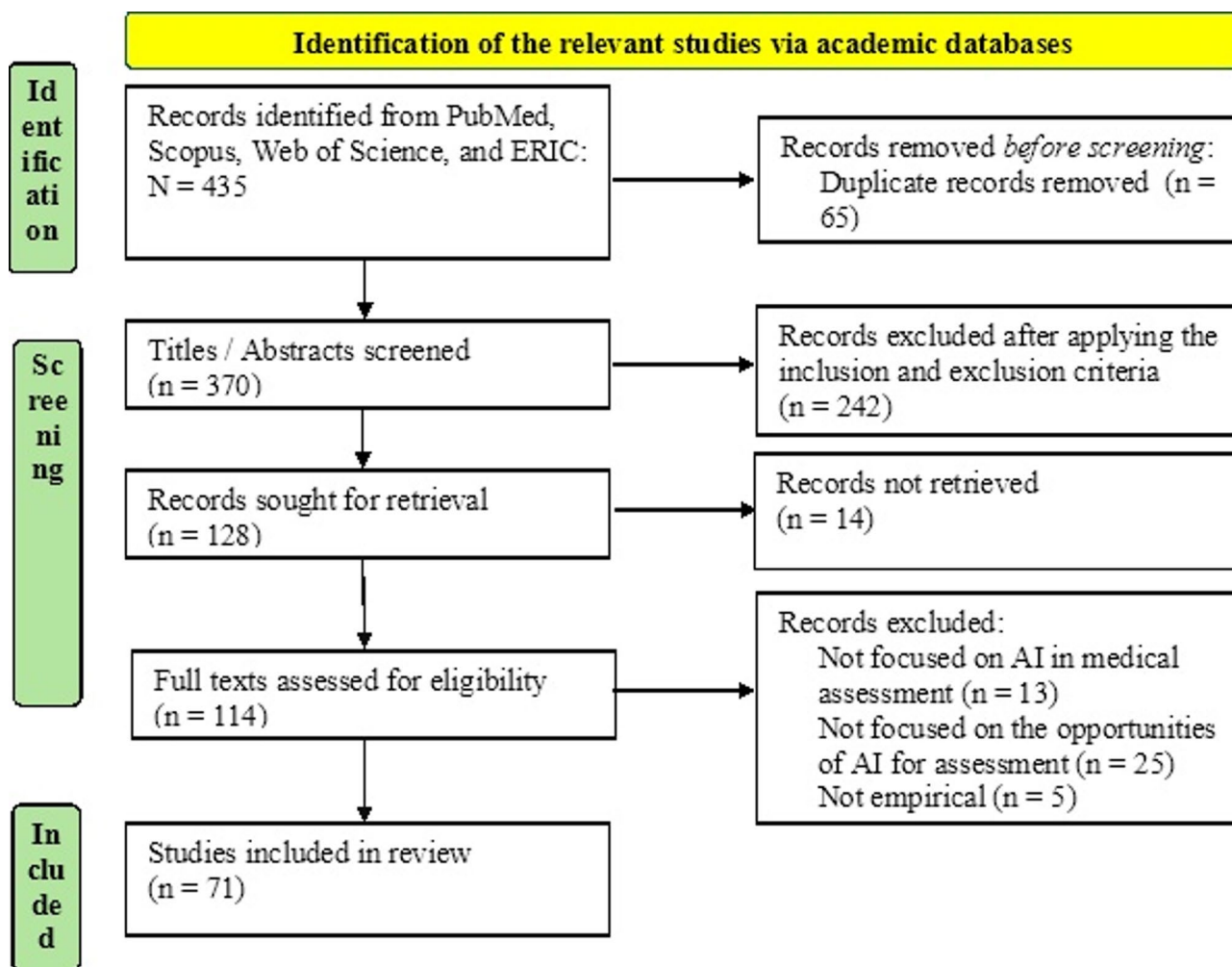


Fig. 1 PRISMA diagram of the systematic review conducted by this study

develop a four-layered model that illustrates the interconnectedness and influence pathways among AI-driven opportunities in medical student assessment.

Search strategy and inclusion criteria

Figure 1 summarises the flow diagram for the systematic review conducted by this study.

As outlined in Fig. 1, a systematic search was conducted across PubMed, Scopus, Web of Science, and ERIC databases using a PCC (Population, Concept, Context) question as follows: What opportunities does AI offer for student assessment in medical education? The search strategy involved combining search terms using Boolean operators. Terms within each category were linked with ‘OR’, and the three categories were linked with ‘AND’. Search terms were applied to the title and abstract fields of each database. The search covered studies published between 2010 and 2025 to capture the recent advancements in AI and its applications in medical education assessment.

Table 1 Search terms used in the systematic review

Category	Search Terms
Population	“medical education” OR “health professions education” OR “medical students”
Concept	“artificial intelligence” OR “AI” OR “learning analytics” OR “machine learning”
Context	“student assessment” OR “evaluation” OR “summative assessment” OR “formative assessment” OR “adaptive testing”

A comprehensive list of search keywords is presented in Table 1.

Screening and data extraction

A total of 435 studies were initially identified. After removing 65 duplicates, 370 records remained. Following a title and abstract screening based on the inclusion and exclusion criteria, 252 studies were excluded, leaving 118 articles for full-text review. Of the 118 articles selected for full-text assessment, 14 were excluded due to unavailability of the full text. The remaining 104 full-text

articles were reviewed in detail against the inclusion and exclusion criteria. As a result, 43 studies were excluded for not meeting the criteria, leaving 71 studies that were included in the final systematic review.

Inclusion criteria were: (1) Peer-reviewed journal articles, (2) Studies focused on AI applications in student assessment, (3) Context: undergraduate, postgraduate, or continuing medical education, (4) Published in English, (5) Empirical, conceptual, or review studies. (6) Published after 2010.

Exclusion criteria included papers that: (1) Focused solely on AI applications outside assessment (e.g., diagnosis training), (2) Did not address medical education or health professions, (3) Were editorials or opinion pieces without substantial data.

Thematic synthesis and categorisation

A reflexive approach to thematic analysis, a method known for its flexibility and rigour in identifying patterns within qualitative data [21], was employed. The researchers' active role throughout the analysis process is emphasised by this approach, and our interpretations and perspectives were acknowledged as central to theme development. The process began with a thorough familiarisation with the literature, after which the identified opportunities for AI in medical student assessment were iteratively coded. Coding was conducted collaboratively by the two researchers, and themes and coding decisions were regularly discussed and refined to ensure consistency and consensus. Through this initial coding process, a total of 22 distinct opportunities were identified. These initial codes were then reviewed and refined, and through a process of constant comparison and synthesis, they were grouped into seven overarching themes, which were termed 'mega-opportunities.' Each mega-opportunity reflects a broader, more abstract theme of AI's potential to address specific challenges within medical education assessment, grounding the findings in the practical realities of the field.

Causal modelling with interpretive structural modelling (ISM)

To explore the significant interdependencies among the seven identified mega-opportunities of AI in medical education assessment, this study employed ISM. ISM is a structured collective intelligence methodology that helps to identify and model complex relationships among components of a system by building a multilevel hierarchical structure [22]. This approach was specifically chosen for its ability to transform expert qualitative judgment into a hierarchical model, which was ideal for this study's exploratory and conceptual nature. Unlike quantitative methods like Structural Equation Modelling, ISM is uniquely suited for mapping complex interdependencies based on collective intelligence rather than large datasets.

Rather than analysing all possible influences among the mega-opportunities, ISM was specifically used to uncover the significant and structurally influential relationships, as determined through expert judgement and consensus during the structural self-interaction matrix (SSIM) development phase. This approach, which aligns with established ISM methodology [22], helps mitigate bias by relying on a transparent and collaborative decision-making process involving domain experts, thereby offering a practical and theoretically grounded roadmap for implementation and future research.

A panel of seven experts from three academic institutions was purposefully selected to participate in the ISM process. The inclusion criteria were: (1) a minimum of five years of professional experience in medical education, (2) direct involvement in the assessment of medical students for at least three academic years, and (3) demonstrated experience with integrating AI in medical education or research, evidenced by at least one peer-reviewed publication, conference presentation, or implementation project. These experts were invited to engage in a structured ISM session during which they responded to pairwise comparison questions generated by the ISM software. The expert judgment was collected through an online, synchronous session where a facilitator guided the panel through the pairwise comparisons. Consensus was achieved in real-time through facilitated discussion, where experts were encouraged to share their rationale for their judgments, especially in cases of initial disagreement. Final decisions for each relationship were then determined by a majority of expert judgments. The core question used to populate the Structural Self-Interaction Matrix (SSIM) was: "Does Mega-opportunity X significantly impact Mega-opportunity Y?"

Following the completion of the SSIM, the ISM process proceeded through standard stages: (1) development of the initial and final reachability matrices, (2) partitioning of levels, and (3) generation of a directed graph (digraph) representing the hierarchical relationships among the mega-opportunities [23]. The final output was a four-layer hierarchical model, clarifying how some mega-opportunities function as foundational drivers while others are outcomes or dependencies within the system. This modelling approach supports better strategic planning and theoretical understanding of AI's role in reshaping medical education assessment.

Critical appraisal and validation

To ensure the reliability and validity of the results, the quality of the studies included in the review was appraised using the Mixed Methods Appraisal Tool (MMAT) [24]. In addition, the thematic synthesis and ISM modelling were reviewed by three experts in the fields of AI, educational assessment, and medical

education, providing feedback to enhance the robustness of the findings.

Findings

The findings of this study are divided into two main sections: (1) Identification and categorisation of AI opportunities in medical education assessment, (2) Modelling the cause-effect interdependencies among those opportunities.

Identification and categorisation of AI opportunities in medical education assessment

The first stage of this study involved conducting a comprehensive literature review to identify and synthesise the diverse opportunities that AI presents for student assessment in medical education. Through an iterative and thematic analysis of the 71 studies included in this review, 22 distinct opportunities were extracted from recent studies, reports, and academic discussions in the field. These opportunities encompass various functions, ranging from personalised learning and competency tracking to simulation-based assessment and curriculum enhancement.

To bring conceptual clarity and manageability to this diverse set of opportunities, a thematic categorisation process was applied. As a result, the 22 opportunities were grouped into seven inclusive categories, hereafter referred to as mega-opportunities. Each mega-opportunity represents a broader functional domain in which AI technologies are transforming the practice of assessment in medical education. These mega-opportunities are as follows: (1) Personalised Learning and Adaptive Assessment, (2) Automated Feedback and Evaluation, (3) Simulation-Based Assessment, (4) Competency-Based and Skills Assessment, (5) Data-Driven Analytics and Curriculum Improvement, (6) Enhancing Focus, Efficiency, and Time Management, (7) Longitudinal Assessment and Development.

Below, each mega-opportunity is defined and the specific opportunities within it are listed in Table 2.

The table above provides a structured synthesis of the diverse ways in which AI is being utilised to enhance student assessment in medical education. Each of the seven mega-opportunities represents a broader functional area where AI demonstrates transformative potential ranging from personalised learning to longitudinal development. By clustering the 22 individual opportunities into these thematic categories (mega-opportunities), the table not only offers conceptual clarity but also highlights overlapping areas of impact (e.g., some opportunities like AI-Assisted Standardised Patient Encounters appear under multiple mega-opportunities). This categorisation serves as the analytical foundation for the next phase of the study, where these mega-opportunities are further

examined for their interrelationships and causal dynamics using ISM. The clarity provided by this framework is particularly critical in a field where AI's applications are rapidly evolving and often fragmented across various studies and contexts.

Modelling the interrelationships among mega-opportunities using ISM

Building on the categorisation of 22 AI-related opportunities into seven mega-opportunities, this section presents the outcome of the ISM-based modelling that explored their cause-effect interdependencies. The aim was to uncover how these mega-opportunities influence one another and to construct a hierarchical framework that reflects their strategic significance in medical education assessment. The resulting model highlights a four-layer structure, illustrating which mega-opportunities serve as foundational drivers and which depend more heavily on upstream developments. This mapping offers a strategic lens/roadmap for educators and researchers to prioritise interventions and understand how advancing one opportunity might trigger progress across others to boost the application of AI in medical education assessment. Figure 2 shows the interdependency model amongst those seven mega-opportunities.

As could be found in Fig. 2, the ISM analysis produced a four-layered hierarchical model that illustrates the significant interdependencies among the seven identified mega-opportunities in AI-driven student assessment within the field of medical education. The model, which should be read from left to right, organises these opportunities based on their strategic positioning, from foundational drivers to those most influenced by the other mega-opportunities.

At the base of the model (Layer 1), two root opportunities are identified: Automated Feedback and Evaluation and Data-Driven Analytics and Curriculum Improvement. These two mega-opportunities are placed within a shared node, indicating a mutual and significant influence on each other. Their foundational role lies in their capacity to generate real-time, data-informed insights that shape and guide the broader assessment ecosystem. The model suggests that these root opportunities exert a strong causal influence on the rest of the system and should be prioritised for early-stage interventions or strategic investment. Specific AI tools - such as learning analytics dashboards that use predictive models to identify at-risk students - offer valuable opportunities for enhancing assessment in medical education.

Moving forward, these foundational opportunities significantly impact the second layer, which includes Simulation-Based Assessment and Longitudinal Assessment and Development. Positioned as intermediate opportunities, these areas rely on feedback mechanisms and data

Table 2 Mega-opportunities of AI for medical student assessment

Mega-Opportunity	Definition	Corresponding Opportunities	References
Personalised Learning and Adaptive Assessment	AI technologies that tailor assessments, learning paths, and feedback to the individual needs, strengths, and weaknesses of students, enabling a more personalised educational experience.	Personalised Learning Pathways and Remediation Predictive Analytics for Student Performance Automated Generation of Assessment Items Medical Students' AI-driven Self-assessment AI for Automated Generation of Personalised Learning Artifacts	[25–27] [28–31] [32–35] [36, 37] [38–41]
Automated Feedback and Evaluation	AI-driven software that provide timely and automated feedback on student performance, helping students identify their strengths and areas for improvement with precision and efficiency.	Automated Feedback on Medical Procedural Skills Automated Scoring of Open-Ended Questions and Essays AI-Driven Personalised Feedback on Documentation Skills	[42–45] [46–49] [50–53]
Simulation-Based Assessment	AI technologies that enhance the simulation experience by offering realistic clinical scenarios and detailed feedback on student performance, including both cognitive and procedural skills.	Simulation-Based Assessment with AI Feedback AI-Enhanced Simulation Fidelity and Realism AI-Assisted Standardised Patient Encounters	[54–59] [56, 58, 59] [60–64]
Competency-Based and Skills Assessment	AI systems designed to track and assess whether students meet specific clinical, academic, non-technical, and professional competencies, ensuring they are prepared for real-world practice.	Analysis of Communication Skills AI-Driven Competency-Based Assessment AI-Driven Analysis of Medical Imaging for Competency Assessment AI-Assisted Standardised Patient Encounters	[65, 66] [67–69] [70–73] [60–64]
Data-Driven Analytics and Curriculum Improvement	AI-driven tools that collect and analyse data from assessments, performances, and other learning activities to help identify trends, monitor student progress, and improve educational curricula.	AI-Enhanced Clinical Reasoning Assessment AI-Driven Analysis of Learning Analytics Data for Curriculum Improvement Analysis of Clinical Performance Data AI-Driven Analysis of Multimodal Data for Holistic Assessment	[52, 62, 74] [75–77] [78–81] [82–85]
Enhancing Focus, Efficiency, and Time Management	AI tools that support students in managing their study schedules, optimising focus, and improving time management, ultimately aiding in more effective goal setting and task execution.	Adaptive Testing AI-Enhanced Clinical Reasoning Assessment AI-Facilitated Feedback on Team-Based Learning AI-Assisted Remote Clinical Skills Assessment	[86–89] [52, 62, 74] [90, 91] [92–94]
Longitudinal Assessment and Development	AI systems that support long-term tracking of students' academic, clinical, and professional development over time, helping to monitor progress toward mastering competencies and achieving career goals.	Personalised Learning Pathways and Remediation AI-Facilitated Continuous Professional Development (CPD) Assessment AI-Assisted Standardised Patient Encounters Automated Generation of Assessment Items	[25–27] [95–98] [60–64] [32–35]

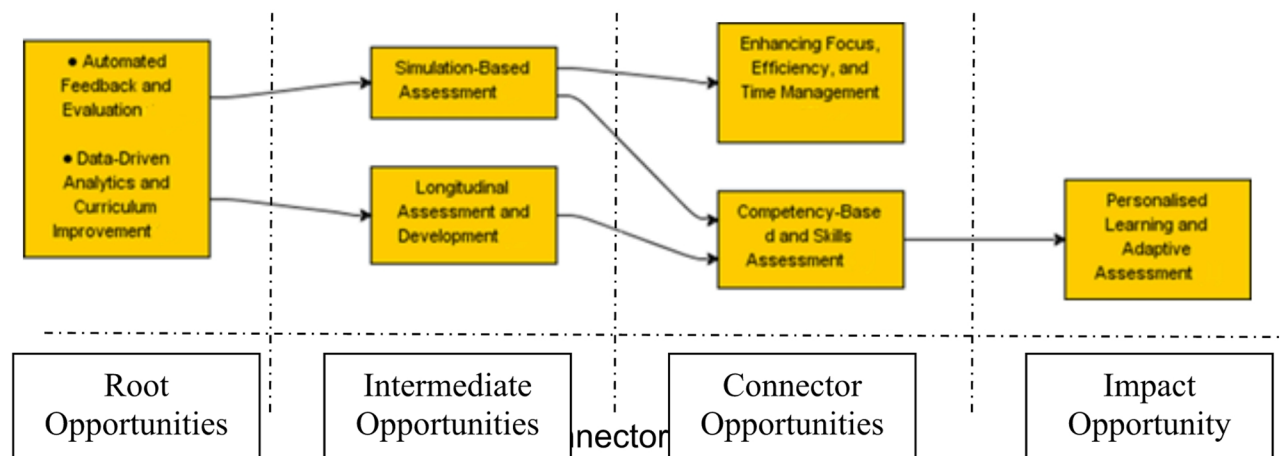


Fig. 2 The interdependency model among the seven mega-opportunities of AI for medical education assessment

analytics to enhance the authenticity, continuity, and effectiveness of assessment over time. These two, in turn, act as bridges to the third layer.

Layer 3 comprises connector opportunities: Competency-Based and Skills Assessment and Enhancing Focus, Efficiency, and Time Management. These mega-opportunities are influenced by the upstream simulation and longitudinal systems and are pivotal in operationalising assessment insights into competency tracking, student workflow management, and broader academic development. They serve as functional links that translate complex assessment data into usable, individualised actions.

Finally, at the right side of the hierarchy (Layer 4), the model positions Personalised Learning and Adaptive Assessment as an impact opportunity. This placement reflects its high level of dependency on all preceding layers. While it represents one of the most promising and student-centred applications of AI, its realisation is contingent upon the successful integration and functioning of all earlier mega-opportunities in the model. It embodies the cumulative effect of enhanced feedback systems, longitudinal tracking, simulation fidelity, and competency analysis.

Overall, this layered model provides a strategic roadmap for educators, researchers, and policymakers by identifying leverage points and sequencing priorities for the integration of AI in medical education assessment. Additional applications of this roadmap are explored in the next section of the manuscript.

Discussion

The ISM-based model developed in this study offers more than a static mapping of AI-driven assessment opportunities; it reveals a strategic and systematic hierarchy of leverage points that should inform both future research and practical implementation of AI tools in medical education. The four-layered structure - root, intermediate, connector, and impact opportunities - demonstrates that the transformative potential of AI in assessment lies not in the technology itself, but in how various innovations are sequenced, integrated, and scaffolded across the educational system.

Critically, the emergence of “Automated Feedback and Evaluation” and “Data-Driven Analytics and Curriculum Improvement” as root opportunities challenges the common tendency in medical education to view AI primarily as a tool for content delivery or assessment efficiency. Instead, this model positions these elements as foundational enablers of educational intelligence, supporting adaptive learning ecosystems rather than isolated interventions. However, while these capabilities offer significant promise - particularly in terms of scalability and consistency - they also raise important concerns about reliability, transparency, and trustworthiness. As

highlighted in recent policy guidance by UK Parliament [10], the accuracy and fairness of AI-generated insights must be carefully scrutinised, particularly when applied to high-stakes educational settings. This underscores the importance of embedding AI within robust learning analytics infrastructures that not only support educational transformation but also uphold ethical and methodological rigour [99].

Moreover, the cascading influence from root to impact opportunities highlights a profound interdependence between formative and summative assessment strategies. Simulation and longitudinal assessment, positioned as intermediary elements, illustrate how contextualised, experiential, and continuous evaluation can effectively bridge these two assessment approaches. By linking real-time formative feedback with longer-term summative outcomes, they contribute to a more cohesive and learner-centred assessment ecosystem. This insight suggests that meaningful adoption of AI in medical education assessment should be developmental rather than transactional to support learners over time and across diverse learning contexts, rather than concentrating solely on isolated test performance.

Particularly striking is the placement of Personalised Learning and Adaptive Assessment at the apex of the four-layered model as the most dependent, rather than initiating, opportunity. While personalised learning is often championed as the primary promise of AI in education, the ISM model reveals that its effectiveness is contingent upon a robust infrastructure and smart implementation of analytics, feedback, simulation, and competency-tracking. This challenges techno-centric discourses that frame personalisation as a direct output of machine learning models (for example in MOOCs), instead arguing for a systems-level, pedagogically grounded approach [100, 101].

From a theoretical standpoint, the model affirms a sociotechnical and ecological view of innovation in medical education assessment. It underscores the importance of moving from a linear adoption of AI tools to a networked, integrated, coherent, and layered logic of educational change, where technology, pedagogy, and assessment evolve in mutual reinforcement. This echoes the growing call in educational AI scholarship for designing AI within pedagogical ecosystems, rather than inserting it into pre-existing frameworks [102–104].

To illustrate the roadmap’s practical application, consider a medical school’s assessment team aiming to strategically integrate AI. Instead of adopting an isolated AI tool for a single purpose, the team could use our roadmap to prioritise their efforts. The model would guide them to first focus on root opportunities, such as piloting Automated Feedback and Evaluation in low-stakes formative quizzes to build faculty trust and validate the

technology's accuracy. Once these foundational elements are established, the team could then progress to intermediate opportunities, integrating AI into Simulation and Longitudinal Assessment to create more cohesive and continuous evaluation systems. Finally, by building on this robust infrastructure, the team would be equipped to effectively implement the impact opportunities, such as Personalised Learning and Adaptive Assessment, which are revealed to be the most dependent and complex to achieve. This strategic, layered approach moves the institution beyond ad-hoc AI adoption, enabling a systematic and pedagogically grounded transformation of its assessment practices.

Finally, the model helps illuminate a major gap in the literature: while many studies report on discrete applications of AI in medical education assessment, few offer a coherent architecture that reveals how those innovations relate to each other and which offer the greatest strategic leverage. This study responds to that gap by not only categorising existing AI opportunities for medical education assessment but also modelling their cause-effect interdependencies, thus offering a roadmap for research initiatives and funding, curricular innovation, and policy prioritisation.

Conclusion

This study has taken a critical step toward addressing the fragmented nature of existing research on the role of AI in student assessment within medical education [105, 106] by offering a structured, multi-layered roadmap. Through a comprehensive literature review, categorisation of 22 AI-driven assessment opportunities, and ISM of their cause-effect interdependencies, this research has generated a strategic framework that not only organises current knowledge but also prioritises high-impact innovation areas. The proposed four-layer model - comprising root, intermediate, connector, and impact opportunities - serves as both a diagnostic tool for understanding the current landscape and a developmental guide for shaping future research and implementation strategies [107, 108].

While our ISM model provides a strategic roadmap for adopting AI in medical education assessment, its practical implementation requires careful consideration of institutional context and ethical principles. A phased piloting approach is recommended, beginning with the foundational 'root' opportunities like automated feedback in low-stakes formative settings to build stakeholder trust and refine institutional workflows. As institutions progress along the roadmap, particular attention must be paid to the ethical, trust, and equity implications, especially where AI is integrated into high-stakes, summative assessments. This involves establishing clear policies for algorithmic transparency, ensuring datasets used for model training are free from bias, and providing robust

mechanisms for human oversight and appeal. Upholding these principles is not just a matter of compliance; it is fundamental to ensuring that AI-driven innovations in medical education are not only effective but also fair, equitable, and trustworthy for all learners. This roadmap, therefore, serves as a guide for both pedagogical innovation and responsible governance, advocating for an implementation strategy that is as thoughtful as it is transformative [109].

Importantly, the roadmap presented here is not intended as a static or prescriptive solution. Rather, it provides a living framework that can evolve with advancements in AI, pedagogical theory, and educational technology. Future studies may refine and expand this model by validating it across different medical education contexts, incorporating diverse stakeholder perspectives (including academics, students and patients), and exploring potential ethical, regulatory, and cultural dimensions [110, 111]. Moreover, institutions and policymakers can adopt this roadmap as a strategic lens to guide investment, curriculum redesign, and faculty development initiatives, ensuring that AI integration into assessment is both impactful and responsible [112]. As such, this study contributes not only to academic scholarship but also to the systemic transformation of assessment practices in medical education.

Despite the valuable contributions of this study, several limitations should be acknowledged. The categorisation and modelling of AI-driven assessment opportunities were based on a purposive sample of expert opinions and existing literature, which, while rich in insight, may not capture the full diversity of global practices or emerging innovations in medical education. Furthermore, the ISM method identifies significant interdependencies but does not quantify their strength or intensity [113], which future studies using complementary methodologies (e.g., Equational Structural Modelling: ESM) could address. Practically, the proposed roadmap offers medical schools and educational technology developers a strategic framework for prioritising AI-driven innovations in assessment, supporting more personalised, scalable, and competency-aligned educational systems. From a research perspective, this study lays the groundwork for longitudinal investigations into the real-world impact of each mega-opportunity on student learning, professional preparedness, and educational equity. Future research should also explore how this roadmap adapts across diverse educational contexts and evolves in response to rapid advancements in AI capabilities.

Authors' contributions

MRZ led the conception and design of the study, conducted the literature review, applied the Interpretive Structural Modelling (ISM) methodology, and drafted the manuscript. MCK critically reviewed the manuscript in multiple rounds, provided substantive feedback and revisions, and contributed to

the refinement and finalisation of the text. Both authors have approved the submitted version of the manuscript and agree to be personally accountable for their contributions and for ensuring the accuracy and integrity of all aspects of the work.

Funding

This is to confirm that the current study did not receive any funding.

Data availability

All data generated or analysed during this study are included in this published article. Additional details or clarifications related to the methodology and data are available from the corresponding author upon reasonable request. The data are not publicly deposited in a third-party repository but can be shared upon request, subject to reasonable data sharing and copyright considerations.

Ethics approval and consent to participate

Not applicable

Competing interests

The authors declare no competing interests.

Received: 19 June 2025 / Accepted: 26 September 2025

Published online: 26 November 2025

References

1. Popenici SA, Kerr S. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*. 2017;12(1):22.
2. Elbounkify I, Berrada I, Mekouar L, Iraqi Y, Bergou EH, Belhabib H, Nail Y, Wardi S. AI-based identification and support of at-risk students: A case study of the Moroccan education system. 2025. ArXiv Preprint ArXiv:2504.07160.
3. Tolsgaard MG, Pusic MV, Sebok-Syer SS, Gin B, Svendsen MB, Syer MD, Brydges R, Cuddy MM, Boscardin CK. The fundamentals of artificial intelligence in medical education research: AMEE guide 156. *Med Teach*. 2023;45(6):565–73.
4. IMARC Group. AI in education market 2024–2033: Share, size, trends, scope, analysis and forecast. 2024; Extracted from <https://www.imarcgroup.com/ai-in-education-market> on 13th April 2025.
5. Holmboe ES, Osman NY, Murphy CM, Kogan JR. The urgency of now: rethinking and improving assessment practices in medical education programs. *Acad Med*. 2023;98(8):537–49.
6. Martinsen SS. Quality learning in undergraduate medical education-Improving teaching and assessment practices. Doctoral theses at NTNU; 2024.
7. Barari N, Rezaei-Zadeh M, Khorasani A, Alami F. Designing and validating educational standards for e-teaching in virtual learning environments (VLEs), based on revised bloom's taxonomy. *Interact Learn Environ*. 2022;30(9):1640–52.
8. Hauer KE, Park YS, Bullock JL, Tekian A. My assessments are biased! Measurement and sociocultural approaches to achieve fairness in assessment in medical education. *Acad Med*. 2023;98(8):516–27.
9. Coates WC. Precision education—a call to action to transform medical education. *Int J Emerg Med*. 2025;18(1):21.
10. Felix J, Webb L. Use of artificial intelligence in education delivery and assessment. Parliamentary Office of Science and Technology; 2024. p. 23.
11. Lin CC, Huang AY, Lu OH. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learn Environ*. 2023;10(1):41.
12. Alqahtani T, Badreldin HA, Alrashed M, Alshaya AI, Alghamdi SS, Bin Saleh K, et al. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Res Social Adm Pharm*. 2023;19(8):1236–42.
13. Mahamuni AJ, Tonpe SS. Enhancing educational assessment with artificial intelligence: Challenges and opportunities. In: 2024 International Conference on Knowledge Engineering and Communication Systems (Ickecs). 2024;1:1–5. IEEE.
14. Desai SV, Burk-Rafel J, Lomis KD, Caverzagie K, Richardson J, O'Brien CL, Andrews J, Heckman K, Henderson D, Prober CG, Pugh CM. Precision education: the future of lifelong learning in medicine. *Acad Med*. 2024;99(4):S14–20.
15. Zarei M, Mamaghani HE, Abbasi A, Hosseini MS. Application of artificial intelligence in medical education: a review of benefits, challenges, and solutions. *Med Clínica Práctica*. 2024;7(2):100422.
16. Goksel N, Bozkurt A. Artificial intelligence in education: current insights and future perspectives. In: Handbook of research on learning in the age of trans-humanism. 2019. p. 224–36. IGI Global.
17. Mir MM, Mir GM, Raina NT, Mir SM, Miskeen E, et al. Application of artificial intelligence in medical education: current scenario and future perspectives. *J Adv Med Educ Professionalism*. 2023;11(3):133.
18. Creswell JW. *Research design: Qualitative, Quantitative, and mixed methods Approaches*. SAGE; 2014.
19. Warfield JN. *An introduction to systems science*. World scientific; 2006.
20. Attri R, Dev N, Sharma V. Interpretive structural modelling (ISM) approach: an overview. *Res J Manage Sci*. 2013;2319(2):1171.
21. Clarke V, Braun V, Commentary. Thematic analysis. *J Posit Psychol*. 2017;12(3):297–8.
22. Warfield JN. Societal systems planning, policy and complexity. *J Cybern*. 1978;8(1):113–5.
23. Rezaei-Zadeh M. An analysis of core entrepreneurial competencies, their interdependencies and their cultivating approaches in virtual education using a collective intelligence methodology, (Doctoral dissertation, University of Limerick). 2014.
24. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Educ Inf*. 2018;34(4):285–91.
25. Ali M, Wahab IB, Huri HZ, Yusoff MS. Personalised learning in higher education for health sciences: a scoping review protocol. *Syst Rev*. 2024;13(1):99.
26. Sunmboye K, Strafford H, Noorestani S, Wilson-Pirie M. Exploring the influence of artificial intelligence integration on personalized learning: a cross-sectional study of undergraduate medical students in the United Kingdom. *BMC Med Educ*. 2025;25(1):570.
27. Mishra M, Dubey V, Hackett TA, Kashyap MK. Artificial intelligence and machine learning in clinical research and patient Remediation. In: *Artificial intelligence and machine learning in healthcare*. Singapore: Springer Nature Singapore. 2023. p. 39–57.
28. Garibaldi BT, Hollon M, Knopp MI, Winkel AF, Burk-Rafel J, Cretta-Weyer HA. How data analytics can be leveraged to enhance graduate clinical skills education. *J Grad Med Educ*. 2025;17(2s):37–41.
29. Colborn K, Brat G, Callcut R. Predictive analytics and artificial intelligence in surgery—opportunities and risks. *JAMA Surg*. 2023;158(4):337–8.
30. Kothinti RR. Artificial intelligence in healthcare: revolutionizing precision medicine, predictive analytics, and ethical considerations in autonomous diagnostics. *World J Adv Res Reviews*. 2024;19(3):3395–406.
31. Manlihot C. Machine learning for predictive analytics in medicine: real opportunity or overblown hype? *Eur Heart J*. 2018;19(7):727–8.
32. Kiyak YS, Kononowicz AA. Using a hybrid of artificial intelligence and template-based method in automatic item generation to create multiple-choice questions in medical education: hybrid AIG. *MedRxiv*. 2024:2024–07.
33. Royal KD, Hedgpeth MW, Jeon T, Colford CM. Automated item generation: the future of medical education assessment. *EMJ Innov*. 2018;2(1):88–93.
34. Qiu Y, Liu C. Capable exam-taker and question-generator: the dual role of generative AI in medical education assessment. *Global Med Educ*. 2025;(0). <https://doi.org/10.1515/gme-2024-0021>.
35. Falcão F, Costa P, Pêgo JM. Feasibility assurance: a review of automatic item generation in medical assessment. *Adv Health Sci Educ Theory Pract*. 2022;27(2):405–25.
36. Khojasteh L, Kafipour R, Pakdel F, Mukundan J. Empowering medical students with AI writing co-pilots: design and validation of AI self-assessment toolkit. *BMC Med Educ*. 2025;25(1):159.
37. Pitrone A, Haddiya I, AI for Education. A generative AI-Powered cognitive tool for medical students' Self-Assessment. *Artificial intelligence in medicine and healthcare*. Apr. 2025;8:67.
38. Fontana A, Sinatti G, Cicchinelli S, Cirella A, Santini S Jr, Caputo V, et al. Automatic generation of learning material for residential students in emergency departments (ED): evaluation of AI-generated flashcards for supporting medical management of gastrointestinal emergencies. *Dig Liver Dis*. 2025;57:S46.
39. Wong DC, Williams S. Artificial intelligence analysis of videos to augment clinical assessment: an overview. *Neural Regen Res*. 2024;19(4):717–8.
40. Cao X, Liang K, Liao KD, Gao T, Ye W, Chen J, Ding Z, Cao J, Rehng JM, Sun J. Medical video generation for disease progression simulation. 2024. ArXiv Preprint ArXiv:2411.11943

41. Do TD, Shafqat UB, Ling E, Sarda N. PAIGE: Examining learning outcomes and experiences with personalized AI-generated educational podcasts. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025. p. 1–12.
42. Fazlollahi AM, Bakhaidar M, Alsayegh A, Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. *JAMA Netw Open*. 2022;5(2):e2149008.
43. Holden MS, Xia S, Lia H, Keri Z, Bell C, Patterson L, et al. Machine learning methods for automated technical skills assessment with instructional feedback in ultrasound-guided interventions. *Int J Comput Assist Radiol Surg*. 2019;14:1993–2003.
44. Cecilio-Fernandes D, Cnossen F, Coster J, Jaarsma AD, Tio RA. The effects of expert and augmented feedback on learning a complex medical skill. *Percept Mot Skills*. 2020;127(4):766–84.
45. Ma R, Kiyasseh D, Laca JA, Kocielnik R, Wong EY, Chu TN, et al. Artificial intelligence-based video feedback to improve novice performance on robotic suturing skills: a pilot study. *J Endourol*. 2024;38(8):884–91.
46. Ariely M, Nazaretsky T, Alexandron G. Machine learning and Hebrew NLP for automated assessment of open-ended questions in biology. *Int J Artif Intell Educ*. 2023;33(1):1–34.
47. Kruidering M, Truongb BB, Endo K, Olvet DM, Fulton TB, Bird JB, et al. Can generative artificial intelligence (AI) reliably score open-ended questions (OEQs) in the assessment of medical knowledge. *J Pharmacol Exp Ther*. 2024;389:224.
48. Quah B, Zheng L, Sng TJ, Yong CW, Islam I. Reliability of chatgpt in automated essay scoring for dental undergraduate examinations. *BMC Med Educ*. 2024;24(1):962.
49. Yenaeng S, Saelee S, Samai W. Automatic medical case study essay scoring by support vector machine and genetic algorithms. *Int J Inf Educ Technol*. 2014;4(2):132.
50. Bongurala AR, Save D, Virmani A, Kashyap R. Transforming health care with artificial intelligence: redefining medical documentation. *Mayo Clin Proc Digit Health*. 2024;2(3):342–7.
51. Lin SY, Shanafelt TD, Asch SM. Reimagining clinical documentation with artificial intelligence. *Mayo Clin Proc*. 2018;9(5):563–5. Elsevier.
52. Schaye V, DiTullio DJ, Sartori DJ, Hauck K, Haller M, Reinstein I, et al. Artificial intelligence based assessment of clinical reasoning documentation: an observational study of the impact of the clinical learning environment on resident documentation quality. *BMC Med Educ*. 2025;25(1):591.
53. Falchetta FS, De Almeida FK, Lemos JC, Goldim JR, Da Costa CA. Automatic documentation of professional health interactions: a systematic review. *Artif Intell Med*. 2023;137:102487.
54. Elendu C, Amaechi DC, Okatta AU, Amaechi EC, Elendu TC, Ezeh CP, et al. The impact of simulation-based training in medical education: a review. *Medicine (Baltimore)*. 2024;103(27):e38813.
55. Kathrine Kollerup N, Johansen SS, Tolsgaard MG, Lønborg Friis M, Skov MB, Van Berkel N. Clinical needs and preferences for AI-based explanations in clinical simulation training. *Behav Inform Technol*. 2025;44(5):954–74.
56. Zidoun Y, Mardi AE. Artificial intelligence (AI)-based simulators versus simulated patients in undergraduate programs: a protocol for a randomized controlled trial. *BMC Med Educ*. 2024;24(1):1260.
57. Wu Y, Hu K, Chen DZ, Wu J. AI-enhanced virtual reality in medicine: A comprehensive survey. 2024. *ArXiv Preprint ArXiv:2402.03093*.
58. Winkler-Schwartz A, Bissonnette V, Mirchi N, Ponnudurai N, Yilmaz R, Ledwos N, et al. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *J Surg Educ*. 2019;76(6):1681–90.
59. Hamilton A. Artificial intelligence and healthcare simulation: the shifting landscape of medical education. *Cureus*. 2024;16(5):e59747. <https://doi.org/10.7759/cureus.59747>.
60. Fukuzawa F, Yanagita Y, Yokokawa D, Uchida S, Yamashita S, Li Y, Shikino K, Tsukamoto T, Noda K, Uehara T, Ikusaka M. Importance of patient history in artificial intelligence-assisted medical diagnosis: comparison study. *JMIR Med Educ*. 2024;10:e52674.
61. Cheng S, Xiao Y, Liu L, Sun X. Comparative outcomes of AI-assisted ChatGPT and face-to-face consultations in infertility patients: a cross-sectional study. *Postgrad Med J*. 2024;100(1189):851–5.
62. Vijayasekaran G, Rajasree RS, Alam P, Suryavanshi A, Ranjanagi V, Mulla UF. A Novel AI-Assisted e-Consult Platform Integrating Deep Learning for Enhanced Healthcare Access and Diagnostic Precision. In 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI). 2025. p. 1103–1108. IEEE.
63. Scheder-Bieschin J, Blümke B, de Buijzer E, Cotte F, Echterdiek F, Nacsa J, et al. Improving emergency department patient-physician conversation through an artificial intelligence symptom-taking tool: mixed methods pilot observational study. *JMIR Form Res*. 2022;6(2):e28199.
64. Stamer T, Steinhäuser J, Flägel K. Artificial intelligence supporting the training of communication skills in the education of health care professions: scoping review. *J Med Internet Res*. 2023;25:e43311.
65. Sridharan K, Sequeira RP. Evaluation of artificial intelligence-generated drug therapy communication skill competencies in medical education. *Br J Clin Pharmacol*. 2024(8). <https://doi.org/10.1111/bcp.16144>.
66. Khan SA, Taiyara J, Zary N, Otaki F. Artificial Intelligence in Narrative Feedback Analysis for Competency-Based Medical Education: A Review. *Stud Health Technol Inform*. 2025:1423–7.
67. Misra SM, Suresh S. Artificial intelligence and objective structured clinical examinations: using ChatGPT to revolutionize clinical skills assessment in medical education. *J Med Educ Curric Dev*. 2024;11:23821205241263475.
68. Cold KM, Agbontaen K, Nielsen AO, Andersen CS, Singh S, Konge L. Artificial intelligence for automatic and objective assessment of competencies in flexible bronchoscopy. *J Thorac Dis*. 2024;16(9):5718.
69. Oyeniji J, Oluwaseyi P. Emerging trends in AI-powered medical imaging: enhancing diagnostic accuracy and treatment decisions. *Int J Enhanced Res Sci Technol Eng*. 2024;13:2319–7463.
70. Salah Eldin W, Kaboudan A. AI-driven medical imaging platform: advancements in image analysis and healthcare diagnosis. *Journal of the ACS Advances in Computer Science*. 2023. <https://doi.org/10.21608/asc.2023.328064>.
71. Cheng CT, Chen CC, Fu CY, Chaou CH, Wu YT, Hsu CP, Chang CC, Chung IF, Hsieh CH, Hsieh MJ, Liao CH. Artificial intelligence-based education assists medical students' interpretation of hip fracture. *Insights into Imaging*. 2020;11:1–8.
72. Lewis S, Bhyat F, Casmod Y, Gani A, Gumede L, Hajat A, et al. Medical imaging and radiation science students' use of artificial intelligence for learning and assessment. *Radiography*. 2024;30:60–6.
73. Strong E, DiGiammarino A, Weng Y, Basaviah P, Hosamani P, Kumar A, Nevins A, Kugler J, Horn J, Chen JH. Performance of ChatGPT on free-response, clinical reasoning exams. *MedRxiv*. 2023:2023–03.
74. Grunhut J, Marques O, Wyatt AT. Needs, challenges, and applications of artificial intelligence in medical education curriculum. *JMIR Med Educ*. 2022;8(2):e35587.
75. Mehta S, Vieira D, Quintero S, Bou Daher D, Duka F, Franca H, et al. Redefining medical education by boosting curriculum with artificial intelligence knowledge. *Journal of Cardiology & Current Research*. 2020;13(5):124–9.
76. Komenda M, Vita M, Vaitis C, Schwarz D, Pokorná A, Zary N, et al. Curriculum mapping with academic analytics in medical and healthcare education. *PLoS ONE*. 2015;10(12):e0143748.
77. Niehoff JH, Kalaitzidis J, Kroeger JR, Schoenbeck D, Borggreffe J, Michael AE. Evaluation of the clinical performance of an AI-based application for the automated analysis of chest X-rays. *Sci Rep*. 2023;13(1):3680.
78. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800–9.
79. Ali M, Rehman S, Cheema E. Impact of artificial intelligence on the academic performance and test anxiety of pharmacy students in objective structured clinical examination: a randomized controlled trial. *Int J Clin Pharm*. 2025;1–8.
80. Sakelaris PG, Novotny KV, Borvick MS, Lagasca GG, Simanton EG. Evaluating the use of artificial intelligence as a study tool for preclinical medical school exams. *J Med Educ Curric Dev*. 2025;12:23821205251320150.
81. Jacobs C. Examining multimodal AI resources in medical education: the role of immersion, motivation, and fidelity in AI narrative learning. *JMIR Med Educ*. 2025;11(1):e72190.
82. Elabd N, Rahman ZM, Abu Alinnin SI, Jahan S, Campos LA, Baltatu OC. Designing personalized multimodal mnemonics with AI: a medical student's implementation tutorial. *JMIR Med Educ*. 2025;11:e67926.
83. Kasa K, Burns D, Goldenberg MG, Selim O, Whyne C, Hardisty M. Multimodal deep learning for assessing surgeon technical skill. *Sensors (Basel)*. 2022;22(19):7328.
84. Shah AM, Yan X, Khan S, Khurram W, Khan QR. A multi-modal approach to predict the strength of doctor-patient relationships. *Multimedia Tools Appl*. 2021;80:23207–40.

85. Burr SA, Gale T, Kisielevska J, Millin P, Pêgo JM, Pinter G, et al. A narrative review of adaptive testing and its application to medical education. *MedEd-Publish*. 2023;13:221.
86. Nizhenkovska I, Afanasenko O, Nizhenkovskiy A. Adaptive testing in pharmaceutical education: strategies and benefits. *Phytotherapy J*. 2024;2:119–24.
87. Saxena R, Salzle KC, Saxena KS. Precision, personalization, and progress: traditional and adaptive assessment in undergraduate medical education. *Innovative Res Thoughts*. 2023;9(4):216–23.
88. Yang TY, Chien TW, Lai FJ. Web-based skin cancer assessment and classification using machine learning and mobile computerized adaptive testing in a Rasch model: development study. *JMIR Med Inform*. 2022;10(3):e33006.
89. Fatima SS, Sheikh NA, Osama A. Authentic assessment in medical education: exploring AI integration and student-as-partners collaboration. *Postgrad Med J*. 2024;100(1190):959–67.
90. Darvishi A, Khosravi H, Sadiq S, Gašević D. Incorporating AI and learning analytics to build trustworthy peer assessment systems. *Br J Educ Technol*. 2022;53(4):844–75.
91. Zoppo G, Marrone F, Pittarello M, Farina M, Uberti A, Demarchi D, et al. AI technology for remote clinical assessment and monitoring. *J Wound Care*. 2020;29(12):692–706.
92. Arslan B, Eyupoglu G, Korkut S, Turkdogan KA, Altinbilek E. The accuracy of AI-assisted chatbots on the annual assessment test for emergency medicine residents. *J Med Surg Public Health*. 2024;3:100070.
93. Pedrett R, Mascagni P, Beldi G, Padoy N, Lavanchy JL. Technical skill assessment in minimally invasive surgery using artificial intelligence: a systematic review. *Surg Endosc*. 2023;37(10):7412–24.
94. Hachoumi N, Eddabbah M, El Adib AR. Health sciences lifelong learning and professional development in the era of artificial intelligence. *Int J Med Inform*. 2023;178:105171.
95. Cohen B, DuBois S, Lynch PA, Swami N, Noftle K, Arensberg MB. Use of an artificial intelligence-driven digital platform for reflective learning to support continuing medical and professional education and opportunities for inter-professional education and equitable access. *Educ Sci*. 2023;13(8):760.
96. Karunathilake IM, de Abrew A. The impact of artificial intelligence in advancing continuing health professions development in Southeast Asia. *South-East Asian J Med Educ*. 2025. <https://doi.org/10.4038/seajme.v18i2.652>.
97. Chau M, Arruzza ES, Spuur K. Effectiveness of ChatGPT-4o in developing continuing professional development plans for graduate radiographers: a descriptive study. *J Educ Eval Health Prof*. 2024;21:34.
98. Strielkowski W, Grebennikova V, Lisovskiy A, Rakhimova G, Vasileva T. AI-driven adaptive learning for sustainable educational transformation. *Sustain Dev*. 2025;33(2):1921–47.
99. Bulathwela S, Pérez-Ortiz M, Holloway C, Cukurova M, Shawe-Taylor J. Artificial intelligence alone will not democratise education: on educational inequality, techno-solutionism and inclusive tools. *Sustainability*. 2024;16(2):781.
100. Selwyn N. Should robots replace teachers? AI and the future of education. Wiley; 2019.
101. Luckin R. Machine learning and human Intelligence. The future of education for the 21st century. UCL institute of education; 2018.
102. Holmes W, Bialik M, Fadel C. Artificial intelligence in education promises and implications for teaching and learning. Center for Curriculum Redesign; 2019.
103. Rezaei-Zadeh M, Mohagheghiaan R, Vahidi-Asl M. Critical meta-analysis of problem-solving serious games: clear signs of pedagogists' disengagement and over-optimistic expectations. *Int J Serious Games*. 2023;10(2):85–113.
104. Turner L, Hashimoto DA, Vasisht S, Schaye V. Demystifying AI: current state and future role in medical education assessment. *Acad Med*. 2023;16:10–97.
105. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med*. 2018;93(8):1107–9.
106. Yudkowsky R, Park YS, Downing SM, editors. *Assessment in health professions education*. New York: Routledge; 2019. p. 26.
107. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. *JMIR Med Educ*. 2019;5(1):e13930.
108. Mastrogiacomi F. Future proofing a ChatGPT-proof portfolio evidence-based formative assessment. *InEdMedia + Innovate Learning. Adv Comput Educ (AAE)*. 2023;176–180.
109. Hosseini Zarrabi H, Rezaei-Zadeh M, Khorasani A. Pedagogy-based solutions for making effective interactions in e-learning environments: reflections on hirumi's three levels of interaction. *Technol Pedagogy Educ*. 2024;33(4):405–20.
110. Knopp MI, Warm EJ, Weber D, Kelleher M, Kinnear B, Schumacher DJ, et al. AI-enabled medical education: threads of change, promising futures, and risky realities across four potential future worlds. *JMIR Med Educ*. 2023;9:e50373.
111. Topol E. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK; 2019.
112. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. *JMIR Med Educ*. 2023;9:e48785.
113. Rezaeizadeh M, Ansari M, Morphi I. A practical guide to the research method: interactive management (IM) and interpretive structural modeling (ISM). Tehran: Jihad Daneshgahi; 2013.
114. Sauerbrei A, Kerasidou A, Lucivero F, Hollowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. *BMC Med Inform Decis Mak*. 2023;23(1):73.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.