University of Central Lancashire UCLan

Central Lancashire Online Knowledge (CLoK)

Title	A Scoping Review of Artificial Intelligence in Medical Education: BEME Guide No. 84
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/50619/
DOI	https://doi.org/10.1080/0142159X.2024.2314198
Date	2024
Citation	Gordon, Morris, Daniel, Michelle, Ajiboye, Aderonke, Uraiby, Hussein, Xu, Nicole Y., Bartlett, Rangana, Hanson, Janice, Haas, Mary, Spadafore, Maxwell et al (2024) A Scoping Review of Artificial Intelligence in Medical Education: BEME Guide No. 84. Medical Teacher, 46 (4). pp. 446-470. ISSN 0142-159X
Creators	Gordon, Morris, Daniel, Michelle, Ajiboye, Aderonke, Uraiby, Hussein, Xu, Nicole Y., Bartlett, Rangana, Hanson, Janice, Haas, Mary, Spadafore, Maxwell, Grafton-Clarke, Ciaran, Gasiea, Rayhan Yousef, Michie, Colin, Corral, Janet, Kwan, Brian, Dolmans, Diana and Thammasitboon, Satid

It is advisable to refer to the publisher's version if you intend to cite from the work. https://doi.org/10.1080/0142159X.2024.2314198

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/

Medical Teacher

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/imte20

A scoping review of artificial intelligence in medical education: BEME Guide No. 84

Morris Gordon, Michelle Daniel, Aderonke Ajiboye, Hussein Uraiby, Nicole Y. Xu, Rangana Bartlett, Janice Hanson, Mary Haas, Maxwell Spadafore, Ciaran Grafton-Clarke, Rayhan Yousef Gasiea, Colin Michie, Janet Corral, Brian Kwan, Diana Dolmans & Satid Thammasitboon

To cite this article: Morris Gordon, Michelle Daniel, Aderonke Ajiboye, Hussein Uraiby, Nicole Y. Xu, Rangana Bartlett, Janice Hanson, Mary Haas, Maxwell Spadafore, Ciaran Grafton-Clarke, Rayhan Yousef Gasiea, Colin Michie, Janet Corral, Brian Kwan, Diana Dolmans & Satid Thammasitboon (29 Feb 2024): A scoping review of artificial intelligence in medical education: BEME Guide No. 84, Medical Teacher, DOI: 10.1080/0142159X.2024.2314198

To link to this article: https://doi.org/10.1080/0142159X.2024.2314198

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

View supplementary material

Published online: 29 Feb 2024.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=imte20

MEDICAL TEACHER
https://doi.org/10.1080/0142159X.2024.2314198

MEDICAL TEACHER
Taylor & Francis
Taylor & Francis Group

BEME GUIDE

OPEN ACCESS Check for updates

A scoping review of artificial intelligence in medical education: BEME Guide No. 84

Morris Gordon[a,b]*, Michelle Daniel[c]*, Aderonke Ajiboye[a], Hussein Uraiby[d], Nicole Y. Xu[c], Rangana Bartlett[e], Janice Hanson[f], Mary Haas[g], Maxwell Spadafore[g], Ciaran Grafton-Clarke[h], Rayhan Yousef Gasiea[b], Colin Michie[a], Janet Corral[i], Brian Kwan[c], Diana Dolmans[j] and Satid Thammasitboon[k]

[a]School of Medicine and Dentistry, University of Central Lancashire, Preston, UK; [b]Blackpool Hospitals NHS Foundation Trust, Blackpool, UK; [c]School of Medicine, University of California, San Diego, SanDiego, CA, USA; [d]Department of Cellular Pathology, University Hospitals of Leicester NHS Trust, Leicester, UK; [e]Department of Cognitive Science, University of California, San Diego, CA, USA; [f]Department of Medicine and Office of Education, School of Medicine, Washington University in Saint Louis, Saint Louis, MO, USA; [g]Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI, USA; [h]University of East Anglia, Norwich, UK; [i]Department of Medicine, University of Nevada Reno, School of Medicine, Reno, NV, USA; [j]School of Health Professions Education, Faculty of Health, Maastricht University, Maastricht, NL, USA; [k]Center for Research, Innovation and Scholarship in Health Professions Education, Baylor College of Medicine, Houston, TX, USA

ABSTRACT

Background: Artificial Intelligence (AI) is rapidly transforming healthcare, and there is a critical need for a nuanced understanding of how AI is reshaping teaching, learning, and educational practice in medical education. This review aimed to map the literature regarding AI applications in medical education, core areas of findings, potential candidates for formal systematic review and gaps for future research.

Methods: This rapid scoping review, conducted over 16 weeks, employed Arksey and O'Malley's framework and adhered to STORIES and BEME guidelines. A systematic and comprehensive search across PubMed/MEDLINE, EMBASE, and MedEdPublish was conducted without date or language restrictions. Publications included in the review spanned undergraduate, graduate, and continuing medical education, encompassing both original studies and perspective pieces. Data were charted by multiple author pairs and synthesized into various thematic maps and charts, ensuring a broad and detailed representation of the current landscape.

Results: The review synthesized 278 publications, with a majority (68%) from North American and European regions. The studies covered diverse AI applications in medical education, such as AI for admissions, teaching, assessment, and clinical reasoning. The review highlighted AI's varied roles, from augmenting traditional educational methods to introducing innovative practices, and underscores the urgent need for ethical guidelines in AI's application in medical education.

Conclusion: The current literature has been charted. The findings underscore the need for ongoing research to explore uncharted areas and address potential risks associated with AI use in medical education. This work serves as a foundational resource for educators, policymakers, and researchers in navigating AI's evolving role in medical education. A framework to support future high utility reporting is proposed, the FACETS framework.

ARTICLE HISTORY
Received 4 December 2023
Accepted 31 January 2024

Background

The use of artificial intelligence (AI) in medicine is expected to have a significant impact on patient care, medical research, and health systems (Weidener and Fischer 2023). The integration of artificial intelligence (AI) and specific tools or techniques to access it, such as natural language processing (NLP), machine learning (ML) and generative pre-trained transformers (GPT), have the potential to transform both medical education content and processes. As technology continues to advance, it is essential to examine the impact of these developments on all aspects of medical education, from admissions to curricula, teaching, and assessment.

Practice points

- What is the extent of evidence on the use of AI (in all forms including NLP, ML and GPT) in medical education and what methods of study have been employed?
- What are the key content themes and areas of focus for these works?
- What, if any, areas are appropriate for more in-depth systematic review?
- What are the gaps in the evidence base currently that should be considered by future researchers?
  To aid consistent and transparent use of terms throughout this review, Table 1 gives some definitions we have employed throughout.

CONTACT Morris Gordon mgordon@uclan.ac.uk School of Medicine and Dentistry, University of Central Lancashire, Preston, UK

Supplemental data for this article can be accessed online at https://doi.org/10.1080/0142159X.2024.2314198.

*Co-first authors.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Whilst the field is not new, the recent application of AI based on machine learning and large language models (LLM), which form the core of AI based chat applications such as ChatGPT by OpenAI has made the technology more widely accessible and easy to use. Masters (2023a) characterised this monumental step as one that could lead to consideration of a pre and post ChatGPT world.

With the rapid expansion of AI, medical educators have become increasingly concerned about the associated risks. A recent piece from a physician defense organization (Graham 2023) warned of the risks of students employing these technologies for their assignments. Other works have cautioned the community concerning AI's potential to undermine the integrity of academic publishing (Masters 2023b; Loh 2023). Educators have also become more excited by AI, as recent landmark works have highlighted the power of AI to enhance assessment (e.g. Schaye et al. 2022), and make educational processes such as admissions and selection more efficient (e.g. Burk-Rafel et al. 2021).

A review in 2021 by Lee J et al. found a diverse range of 22 published studies on what and how to teach AI within undergraduate medical education (UME), but the wider fields of graduate medical education (GME) and continuing professional development (CPD) were outside of scope. Publications and interest in AI has increased exponentially in the last 12 months in response to the launch of ChatGPT, indicating the need for a new review of the evidence.

This review aims to provide insights into the current state of AI applications and challenges within the full continuum of medical education. We chose a scoping review methodology, mapping out the landscape of existing literature, to delineate the evidence base, classify methodologies, and highlight thematic content areas ripe for systematic investigation. We aim to provide the path forward, guiding AI integration into medical education and ensuring that healthcare professionals are adeptly equipped for the evolving and increasingly complex healthcare milieu.

Methods

This scoping review was conducted in a rapid timeframe – completed within 16 weeks of its inception. Despite the expedited process, we ensured systematicity, and maintained integrity and methodological precision, in line with prior publications (Gordon et al. 2020; Daniel et al. 2021). Our methodology was guided by Arksey and O'Malley's (2005) scoping review framework, which was refined by Levac et al. (2010). In terms of reporting, we adhered to the STORIES statement (STructured apprOach to the Reporting In healthcare education of Evidence Synthesis) (Gordon and Gibbs 2014) and BEME guidance (Hammick et al. 2010). We also conformed to the PRISMA-ScR reporting standards (Tricco et al. 2018), drawing on support from Peters et al. (2020, 2022) to offset the absence of specialized reporting standards for scoping reviews in health professions education. A protocol is available at the following link (https://clok.uclan.ac.uk/49900/).

The five stages of a scoping review were followed (Arksey and O'Malley 2005):

Stage 1: identifying the research aims/questions (as stated above)

Stage 2: identifying relevant studies

Stage 3: study selection

Stage 4: charting the data

Stage 5: collating, summarizing and reporting the results

The sixth stage recommended by Levac et al. (2010), expert consultation, was not performed, as AI (BK, JC, MS) and educational experts (CM, DD, JC, JH, MD, MG, MH, ST) on the author team fulfilled this role.

Stage 2: Search strategy

In conjunction with an experienced information specialist, we conducted a search of PubMed/MEDLINE, EMBASE, and MedEdPublish without restrictions on date or language. We

Table 1. AI terminology.

Artificial Intelligence (AI)	AI is a *concept*, the most general of the terms, that spans any process that involves a machine acting 'intelligent.' Intelligence is most often defined as 'human-like' in its ability to make decisions, learn from mistakes, generate insights, or understand language (Coppin 2004).
Machine Learning (ML)	ML is a *method* and *discipline*. ML involves the specific mathematical and computational structures which produce computer programs/algorithms that can make decisions given input data. ML is most frequently the way in which we achieve (a semblance of) AI (Gopinath and Churiwala 2019).
Traditional Machine Learning	Traditional ML involves extensive human-engineered preprocessing of input data to extract salient 'features' which are then piped into a mathematical model/equation to produce decisions/results. Examples: logistic regression, support vector machines, decision trees, K-nearest neighbors, gradient boosted machines (Gopinath and Churiwala 2019).
Data Mining (DM)	DM often borrows basic statistics or algorithms from traditional ML (for example, logistic regression or k-nearest neighbors), but instead of trying to teach a machine to make a decision, it instead attempts to sort through large volumes of data to allow humans to make insights and better understand the data itself. Whereas ML is concerned with assessing the ability of machines to make *prospective* decisions on heretofore-unseen data, DM instead tries to glean insights from full, known sets of previously collected data (Islam et al. 2018).
Deep Learning (DL)	Deep ML involves relatively little human-engineered preprocessing of input data, which is often input in raw or near-raw form. In DL, the model itself creates its own internal representations of features, as represented by the connection strength between mathematical 'neurons' within the artificial neural network. Deep neural networks require significant time/ energy/ money/ data investments to initially train. However, these 'pre-trained' models can be 'customized' to new domains, such as MedEd, cheaply (Naylor 2018).
Natural Language Processing (NLP)	NLP is any type of computational or mathematical approach that deals with natural human (written or spoken) language. NLP is almost always paired with one of the approaches above, which is often why you see it written with a slash (e.g. NLP/ML). ChatGPT, for example, is NLP/ML (specifically deep learning), because it is a deep machine-learning artificial neural network which processes natural language (Iroju and Olaleke 2015).

used a comprehensive search strategy with both controlled vocabulary (e.g. MeSH terms) and keywords. The search terms were based on keywords related to AI, ML, NLP, GPT, and medical education. The search strategy evolved through iterative pilot searches to refine the use of complex terminology. The initial search in July 2023 yielded 1,598 publications. Reference list screening identified key publications not captured, prompting further refinement. The final, more inclusive search was executed in August 2023. The final complete search strategy is detailed in Supplementary Appendix 1.

We reviewed the bibliographies of included studies for any missed articles.

Stage 3: Study selection

The screening process was conducted by pairs of authors (AA, HU, MD, MG, RG) who independently reviewed titles, abstracts, and full texts. Discrepancies were resolved by consensus or with a third author's input. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram was constructed to outline the publication selection steps.

Inclusion criteria

- Publications that described the use of AI, ML, NLP, or GPT in medical education
- Publications that discussed AI, its impact, attitude, knowledge or view, regardless of whether they deployed AI
- Publications in undergraduate medical education (UME), graduate medical education (GME) or continuing professional development
- Publications with medical students, residents, fellows, or physicians. If discussing multi-professional learner groups, medics comprised a significant proportion.
- Publications focused on all relevant areas, including curriculum, teaching and learning, or assessment
- Publications of all methodological forms
- Publications in any language

Exclusion criteria

- Publications that discussed the use of AI for diagnostic, clinical, healthcare organizational systems or governance roles in healthcare
- Publications that employed AI exclusively as a research tool in health professions education
- Publications employing AI to write their work without focusing on AI within their work

Stage 4: Charting the data

The approach taken was amended for each of two clear bodies of literature that were included.

The first category were articles and Innovations. These include scholarly reports of original studies of the proposed or actual employment of AI or related issues in medical education.

The second category were perspectives publications. These include scholarly pieces that present a point of view on AI issues in the field of medical education. They are not original study reports but rather informed opinions that draw on the authors' expertise, experience, or review of the existing literature.

Articles and innovations

A team of eight authors (AA, BK, CJC, HU, MH, MS, NX, RB), working in pairs, charted the data. They used a predefined data extraction form, which was piloted and refined, with the following data extracted from each study:

- Study identifiers (authors, title, publication date, journal)
- Study characteristics (stage of education, country of origin, medical specialty, study aims, study design)
- AI application or focus of use (e.g. admissions and selection, curriculum, assessment, etc.)
- AI method (e.g. was AI treated as a general concept, or were specific methods such as DM, traditional ML, DL, and/or NLP used?)
- AI use cases
- Name of language model and software programs, if applicable
- Rationale for AI use
- Categorization per the SAMR technology integration framework (substitution, augmentation, modification, redefinition)
- Description of AI implementation
- Summary of results and Kirkpatrick's outcomes, if applicable
- Implications for future practice, policy or research

Perspectives

A team of four authors (JC, CM, JH, ST), working in two independent pairs, charted not only similar metadata from publications in the Perspective category but also additional details such as the underlying rationale for using AI, its applications, frameworks, and recommended topics for curriculum and research. They further considered ethical issues and the constraints of these applications.

Stage 5: Collating, summarizing and reporting the results

Articles and innovations

Utilizing data from the extraction forms, all authors collated the data into a number of tables and figures for easy visualization, to provide a map of the current evidence base. After charting, we produced a narrative account of our findings that considers the extent and range of developments included in the review, as well as the outcomes assessed. We identified areas where a paucity of research exists. We also suggested areas for future primary and secondary studies (i.e. systematic reviews).

Perspectives

Upon familiarizing itself with the data, the team developed a systematic approach for synthesis. Publications were first organized into two main categories reflecting the central themes of their perspectives: 'Importance of Integrating AI into Medical Education' and 'Potential Application into Medical Education.'

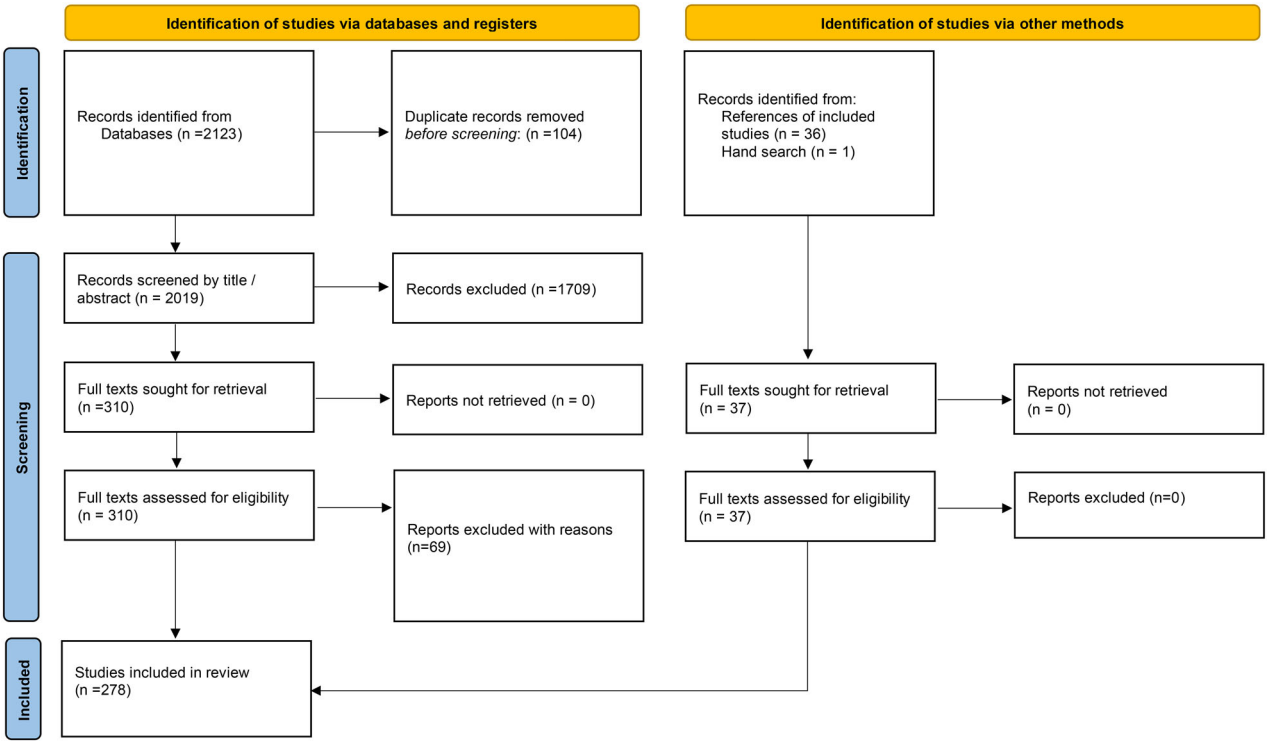

Figure 1. Prisma flow diagram.

The sub-team then applied a utility-focused approach to the latter category, highlighting the practical value of each publication for readers, termed 'Contributions.' These contributions are pivotal for making informed decisions, designing educational programs, or evaluating AI applications in medical education, despite the fact that most have yet to be studied or evaluated. The Contributions are delineated as follows: 1. Suggested Educational Applications and Developments, 2. Educational Frameworks, 3. Recommendations for Curriculum Content, Competencies, Tools and Resources, and 4. Ethics Considerations.

At each step, team members independently categorized the publications, followed by discussions to reach a consensus. We produced a narrative synopsis of our findings as 'Brief Summary' and 'Key Points' to aid reader comprehension and application.

Results

Overview

The database search yielded 2,123 publications, with 2,019 remaining after duplicate removal. Title and abstract screening narrowed the pool to 310 publications, with a high inter-rater reliability ($\kappa = 0.86$). Full-text review further excluded 70 publications, primarily for using AI solely as a research tool or lacking focus on AI, leaving 241 for inclusion. Hand-searching references added 37 publications, with 278 publications ultimately included in the review. The PRISMA diagram is shown in Figure 1. Full appendix tables with extracted data for all publications have been uploaded to a repository and can be accessed following this link (https://clok.uclan.ac.uk/49919/). An infographic contained in Figure 2 gives much summary detail around the included studies.

The geographic distribution of studies was predominantly Anglo-North American ($n = 138$, 49.6%), followed by European ($n = 51$, 18.4%), East Asian ($n = 22$, 7.9%), Middle Eastern ($n = 22$, 7.6%), and South Asian ($n = 17$, 6.1%) contributions. Other regions produced only a handful of studies and no studies were identified from Africa. The earliest publication on AI in medical education was in 1992. The temporal trend showed a surge in publications from 2018 onwards, with 11 in 2018, 14 in 2019, 18 in 2020, 49 in 2021, 57 in 2022, and 114 as of August 2023 (see Figure 2a,b and Supplementary Appendices 1 (columns E and I) and 2).

Most included publications fall under the category of articles or innovations ($n = 191$, 68.7%), as defined above, with the remainder being perspectives ($n = 87$, 31.3%). Undergraduate medical education (UME) was the focus of 48.6% ($n = 135$), graduate medical education (GME) 22.3% ($n = 62$), and continuing professional development (CPD) 2.5% ($n = 7$). The remaining publications spanned multiple levels (UME, GME, or CPD) ($n = 74$, 26.6%) (see Figure 2c,d and Supplementary Appendix 2 (column H) and 2).

Six basic science disciplines and 24 clinical specialties were represented in the included publications. Anatomy and Physiology were the most common basic science disciplines ($n = 4$, 1.4%, each), whereas radiology ($n = 31$, 11.2%) and surgery ($n = 24$, 8.7%) were the most common clinical specialties, followed by ophthalmology ($n = 8$, 2.9%) and neurosurgery ($n = 6$, 2.2%). One hundred fifty-seven studies (56.5%) did not specify a discipline or specialty (see Figure 2e and Supplementary Appendices 2 (column J) and 2).

Articles and innovations included in the review

This review encompasses a wide range of articles and innovations, systematically cataloged as A1–A191 in Supplementary Table 2, facilitating cross-reference throughout this review. Of these studies, seven (3.6%) used AI for admissions and selection (A1–A7), 34 (17.8%) focused on teaching about AI or AI-augmented instruction (A8–A41), 50

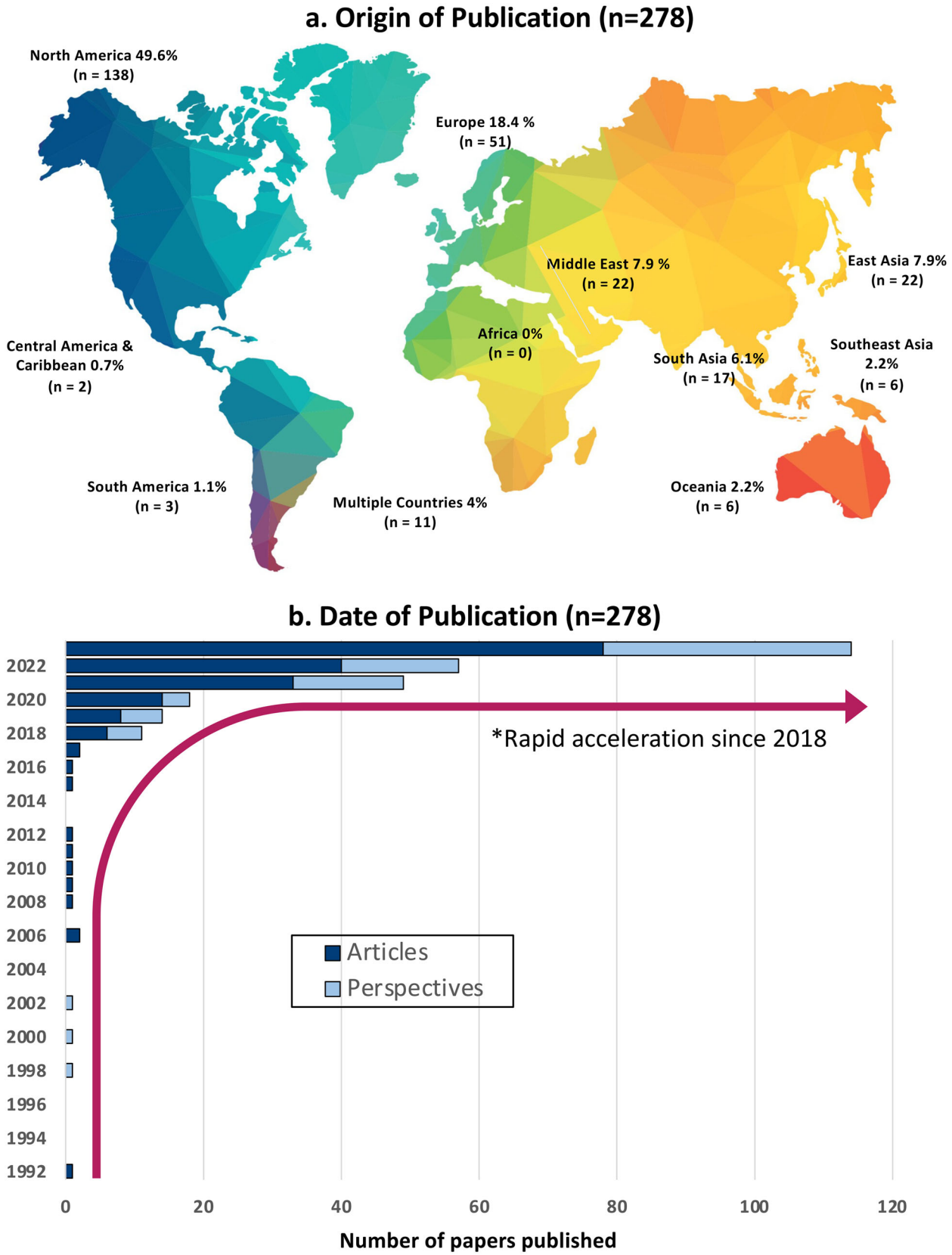

Figure 2. Infographic summarising results.

(26.2%) used AI in assessment (A42–A91), thirteen (6.8%) used AI to teach or assess clinical reasoning (A92–A104), and four (2.1%) used AI for automation of case and procedure logs (A105–A108). A substantial portion ($n = 51$, 26.7%) investigated knowledge or attitudes about AI in medicine (A109–A159), and the performance of large language models (LLMs) on medical questions or exams was evaluated in 32 (16.8%) of the studies (A160–A191. (see Figure 2f and Supplementary Appendix 2, column M).

The technical application of AI varied. In 65 (34%) of the studies, AI was treated as a general concept without actual computation. The remainder of studies engaged in various computations: data mining in 4 (2.1%), traditional ML in 41 (21.5%) and DL in 81 (42.4%). NLP was used in 78 articles

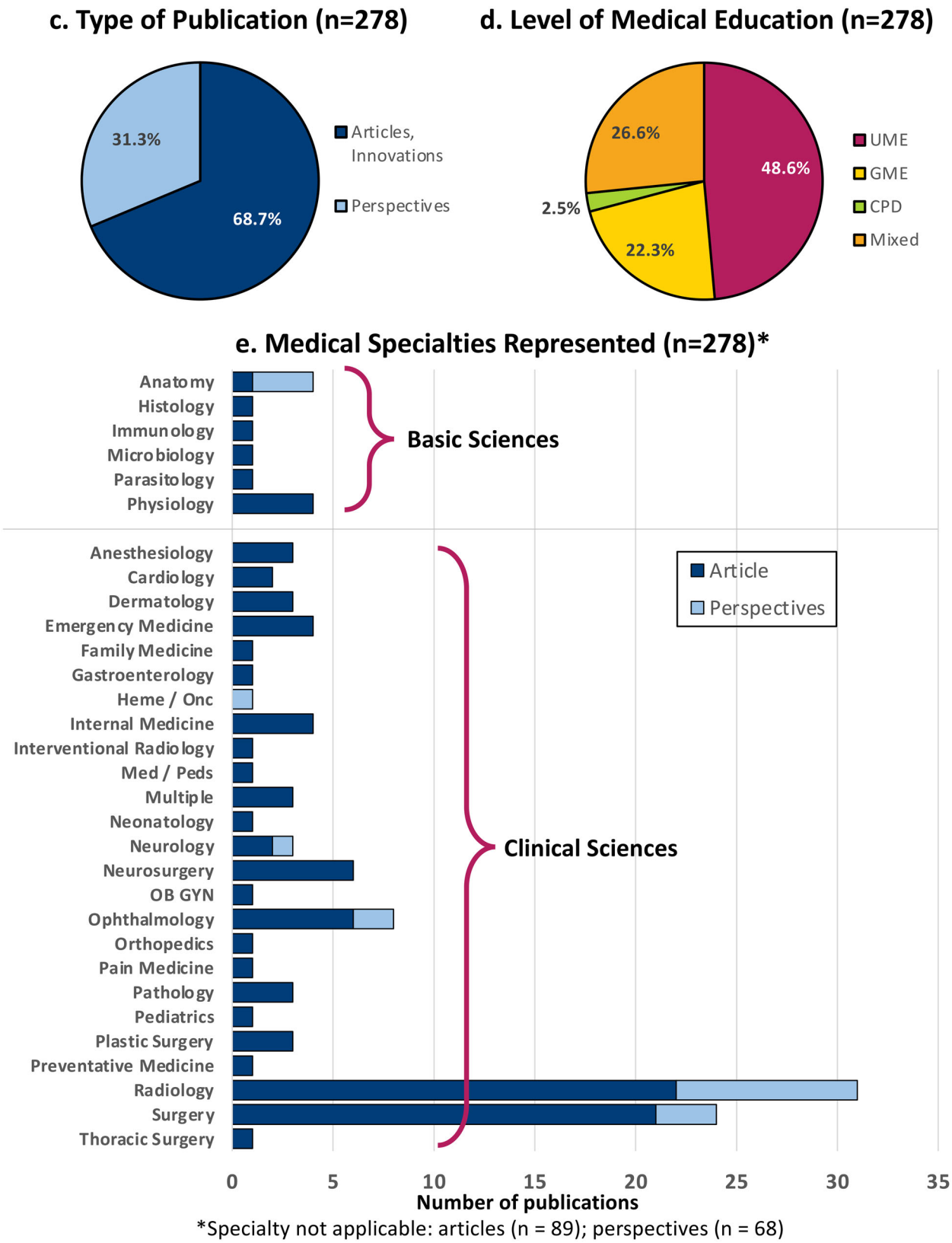

Figure 2. Continued

(40.8%). (See Figure 2g,h, and Supplementary Appendix 2, columns N and O. Also refer to Table 1 for AI terminology.)

Employing the SAMR framework to evaluate four levels of technology integration, this review found that AI primarily served as a substitute in 36 articles (18.8%), enhanced existing processes through augmentation in 34 (17.8%), contributed to task redesign or modification in 10 (5.2%), and facilitated an entirely novel, previously unconceivable task or redefinition in 11 (5.8%). For a substantial number of articles (100, 52.4%) the SAMR framework did not apply (see Figure 2i and Supplementary Appendix 2).

The review also identified a spectrum of AI applications, described as 'use cases,' which included data analytics ($n = 9$), predictive models ($n = 17$), performance analytics ($n = 19$),

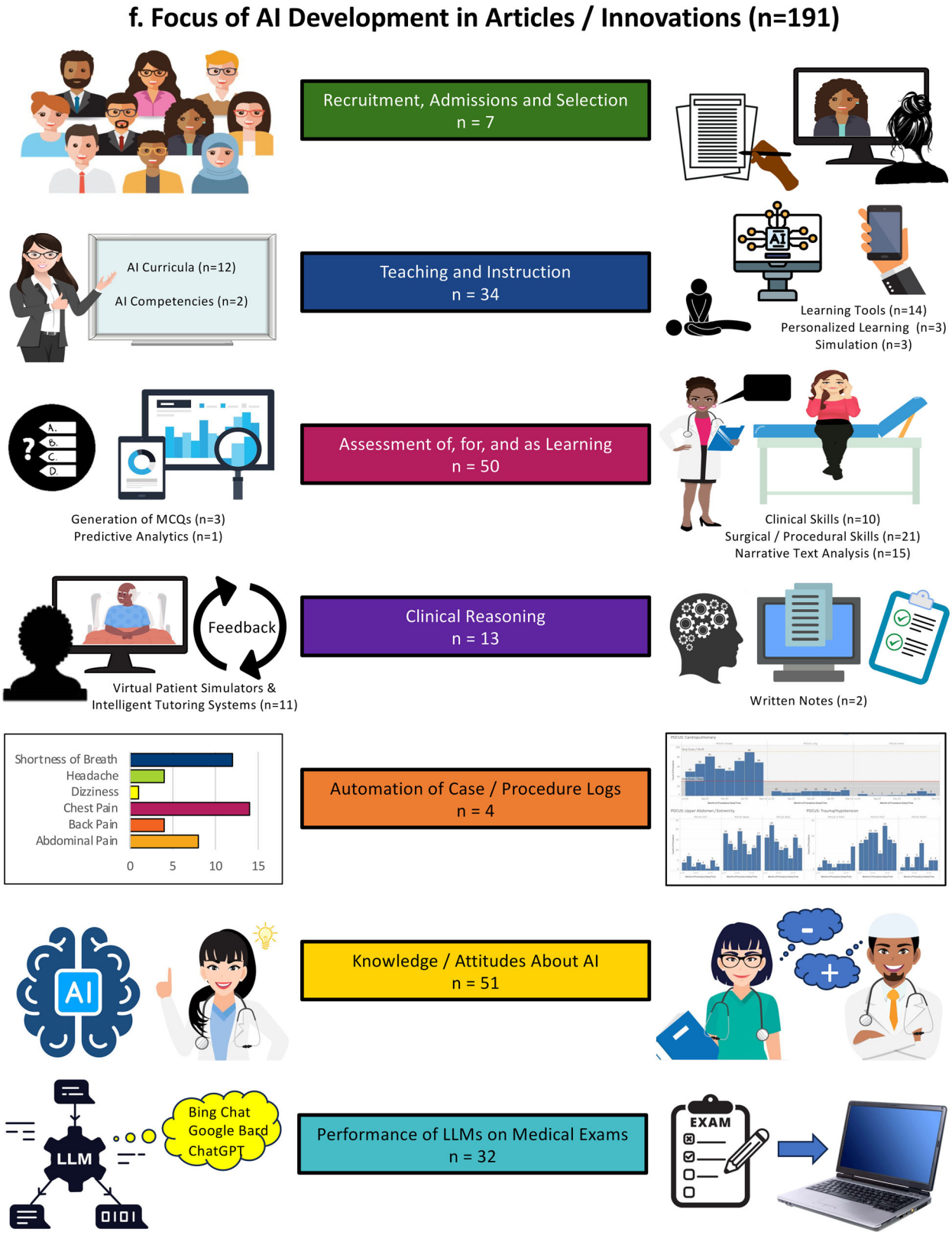

Figure 2. Continued

sentiment analysis or bias evaluation ($n = 7$), virtual patient simulators (VPSs) ($n = 11$), personalized learning platforms or intelligent tutoring systems ($n = 6$), clinical guidance for trainees ($n = 10$), procedural guidance for trainees ($n = 5$), ChatBots ($n = 2$), data labeling ($n = 2$), content generation ($n = 5$), and summative assessment completion ($n = 32$) (see Figure 2j and Supplementary Appendix 2, column P).

AI use in admissions and selection for medical school and residency

In one study (A7), an AI Chatbot was used during recruitment to host a virtual question and answer session. In another study (A5), sentiment analysis was used to detect gender bias in letters of reference for a surgery residency over three decades. This study demonstrated that while

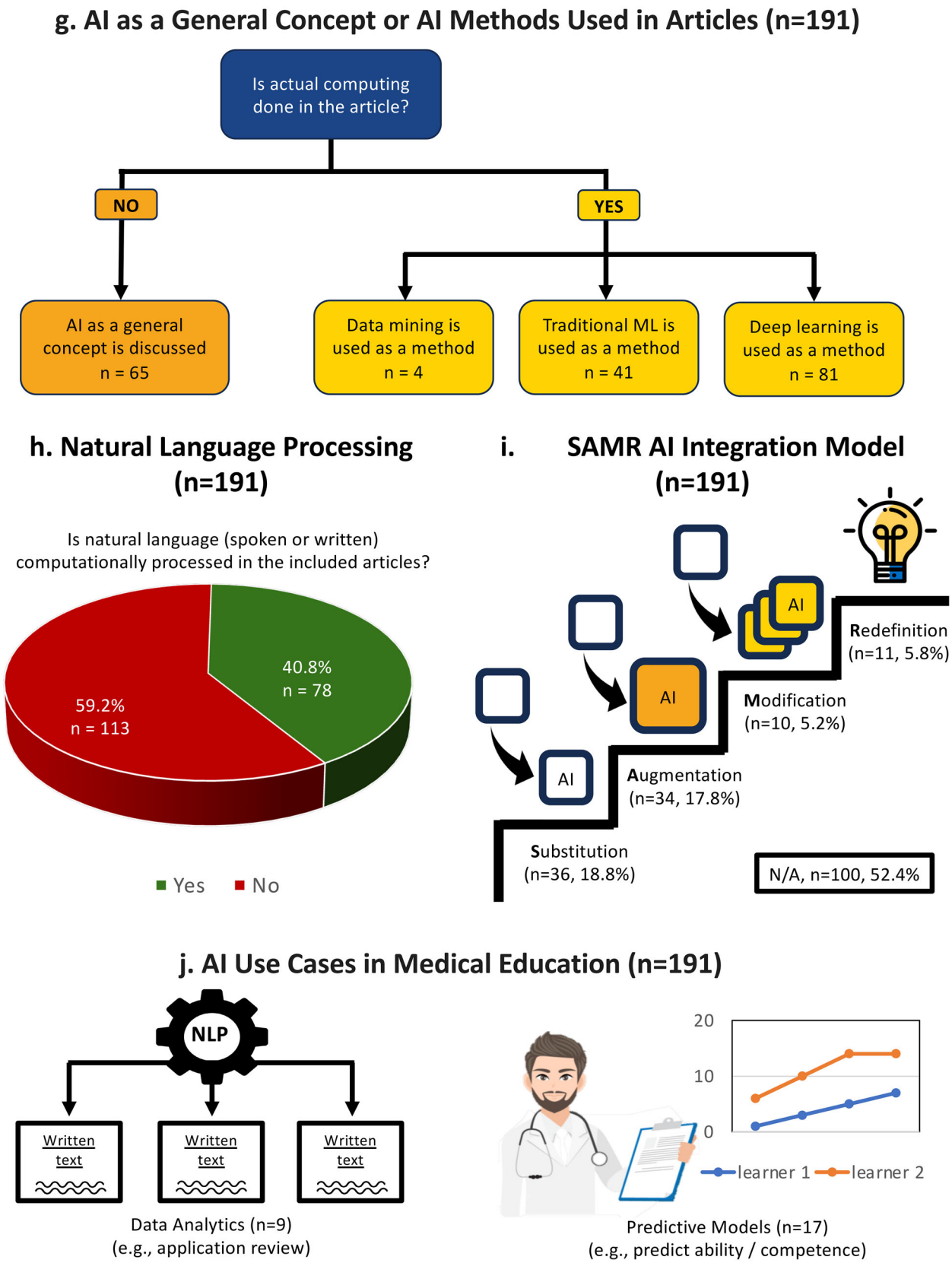

Figure 2. Continued

gender bias has decreased over time, letters continue to be positively biased towards men, impacting selection. Another study (A3) compared faculty scoring of medical student performance evaluations (MSPEs) to a machine learning (ML) model's scores. The ML model was unable to successfully rank MSPEs using commercially available sentiment analysis programs, largely because all MSPEs were highly positive. A bespoke model may be developed in the future that can better replicate faculty scores.

In one study (A4), ML was used to develop a predictive model to identify candidates likely to be ranked and matched. The predictive accuracy for ranking was outstanding, with an area under the receiver operating characteristic (AUROC) of 0.925. In three studies (2 in UME (A2, A6)

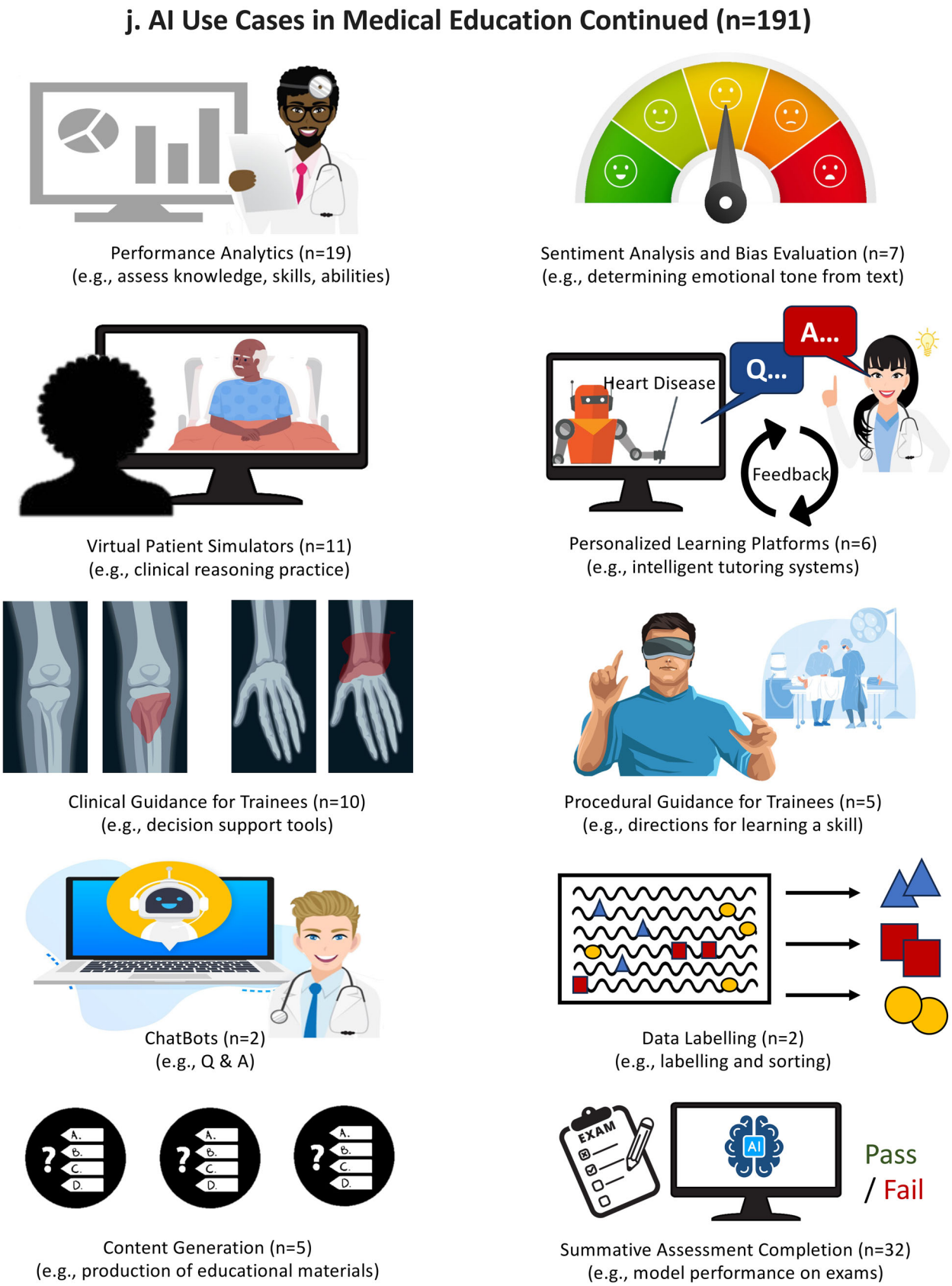

Figure 2. Continued

and 1 in GME (A1)), ML was used to screen applications. The ML models were trained on retrospective data sets, then validated on retrospective or prospective data sets. The models demonstrated impressive accuracy with AUCs of 0.925–0.95, and hold promise in the future for partially automating holistic review of large numbers of applications. When applied in a restorative fashion that augments human review, ML models can be used to mitigate bias and identify applicants who may otherwise have been screened out and not offered an interview (A1). With

further validation and refinement, ML models have the potential to make application screening more efficient and equitable, and the ranking process more accurate.

Teaching about AI and AI-augmented instruction

A total of thirty–three articles addressed teaching about or instruction with AI (A8–41). Of these, 12 focused on AI as a learning tool, 11 on curriculum development, three on AI use for personalized education, five on the use of AI in medical simulation, and two on AI educational competencies.

The majority of articles (7 out of 12) focused on AI as a learning tool, training users to identify findings on images, including histopathologic identification of glomerulopathies (A22), neuroanatomic structures (A35), general tissue types (A31), and melanomas (A25). A few articles emphasized identification of clinical pathology (e.g. diabetic retinopathy (A26, A30) and hip fractures on plain radiographs (A24)). One article described AI–powered, real–time instruction on acquiring optimal echocardiographic views with point–of–care ultrasound (A32) and another trained students to develop AI algorithms for sonographic identification of pulmonary structures (A33). Another set of articles used the interactive features of chatbot-style tools for anatomy instruction (A29), delivering difficult news (A34), and LLM-based summaries of surgical procedures (A28).

Of the eleven articles that addressed AI curriculum development, a majority ($n = 6$) described the development of instructional modules. Four emphasized fundamental concepts in AI and ML for UME learners (A11, A13, A16, and A19). The GME/CPD studies largely focused on radiologists and included AI workshops (A14, A17), a deep-learning curriculum (A18), a data-science elective (A20) and a medical DM course (A21). One article described an introductory AI course for practicing physicians (A12).

Three articles explored ML for personalized learning, including an NLP algorithm to support trainee evaluation of geriatric patients through just-in-time learning (A36), a ML-powered adaptive learning system for preclinical students (A37), and a DL model for patient triage in ophthalmology (A38).

Five articles examined AI in medical simulation, covering topics like the interview of standardized patients about the use of robotic standardized patients (A39), an AI-assessed cognitive load in simulation through ECG and galvanic skin response (A40), a DL model for simulation outcome evaluation of recorded simulation sessions (A41), a ML-based multimedia simulation of an individual nephron (A23), and generative AI for creating images of dermatologic lesions (A27).

Lastly, two articles delved into educational competencies in AI, one developing a tool for assessing medical student readiness for AI use in clinical settings (A9), and the other examining the prevalence of AI competencies in GME and CPD settings in Germany (A10).

AI use in assessment

Of the 50 articles related to AI use in assessment, a focus on psychomotor skills was prevalent, with 21 articles examining this area: 18 on surgical skills and 3 on procedural skills, with the latter group focusing exclusively on intubation techniques (A71–73). Surgical skills assessments predominantly focused on minimally invasive and robotic surgery within general surgery (A74–A83, A85, A87–A88, A91), with a handful addressing neurosurgical (A84, A86, A89) and pediatric surgery (A90). Fifteen articles focused on narrative text analysis, 10 on clinical skills, 3 on generation of multiple choice questions (MCQs), and 1 on predictive analytics.

A notable trend involved leveraging ML as an alternative to human observation and feedback (A74, A76–A78, A83, A86, A88, A90). Many utilized virtual trainers or simulators to facilitate assessment in simulated environments. A particularly innovative approach involved cloud-based, AI-powered video analytics to monitor and assess users' performance in minimally invasive surgery (A81). A few studies described the development of algorithms capable of discerning expertise levels or predicting surgical skill (A74–A76, A85, A86) or learning curves (A80, A84, A87, A91) based on visual or kinesthetic inputs. One study developed a ML model, integrating EEG data analysis of participants, to distinguish expert from novice surgeons (A89). These studies underscored the potential of AI to mitigate the resource intensity and inherent bias of traditional assessments.

Narrative text assessment is another area where AI is gaining ground, with sentiment analysis and NLP being common tools for extracting insights from textual data. This application of AI promises an efficient and automated analysis of extensive narrative information for enriching and contextualizing quantitative assessments such as ratings. Faculty's narrative evaluations of residents (A55, A57, A62-A64, A67-A68, A69) and clerkship students (A55, A56, A65) were areas where AI was deployed to inform clinical competency committees (A55), gain insights into competency development (A57, A68-A69), generate quality ratings of faculty feedback (A64, A67), facilitate objective interpretation of faculty evaluations (A65), identify learners at risk (A68), and identify biases within written student evaluations (A56, A66). NLP was further used to parse through reflective writings of students (A59), assess professionalism in short answer responses (A61), identify patterns suggestive of professionalism lapses in evaluative feedback among faculty (A60), and evaluate physician performance as it related to professionalism and patient safety (A58).

Among studies of clinical skills assessment, seven out of ten focused on UME applications. Here, VPSs were a mainstay, reducing logistical demands (e.g. time and resource) of traditional assessments, and enabling remote examinations (A42, A48–A49, A51). AI's role extended to virtual Objective Structured Clinical Examinations (OSCEs) (A42, A49) and automated grading of OSCEs (A43–A44, A50) or both (A46, A48, A51). In GME, there were studies describing the use of a virtual provider simulator for resident training in communication skills, with automated assessment and feedback (A47) and NLP for assessment of critical care trainees' oral case presentations (A45).

Three studies delved into AI-generated MCQs, harnessing DL and NLP to create assessment tools and study aids (A52–A54). These studies emphasized AI's capability to streamline question development, substituting intense human efforts, though still requiring human verification to ensure quality. The benefits of AI included speed, and ChatGPT emerged as the preferred tool. Study results were

mixed, however, with some skepticism of AI's readiness to fully take over the MCQ creation process (A52–54).

The final point to note is that predictive analytics played a notable role in evaluating surgical skills across multiple studies. However, only one study (A70) primarily focused on this aspect. This particular study investigated the prediction of medical student performance on high-stakes exams through ML. This illustrates a nascent but promising application of AI in predicting academic success in high-stakes environments.

AI to teach and assess clinical reasoning

Thirteen articles examining AI's role in teaching and assessing clinical reasoning reveal a promising frontier. Eight studies focused on the fusion of AI-powered VPSs and intelligent tutoring systems (ITSs) (A92, A95–A98, A101–A102, A104). VPSs simulate authentic patient interactions, where learners ask questions and receive answers from an avatar or *via* text, offering dynamic environments for learners to hone their reasoning through varied clinical scenarios. This allows learners opportunities to practice various components of the clinical reasoning process, from information gathering and hypothesis generation, to problem representation and diagnostic justification.

The ITSs complemented this by delivering personalized learning and real-time feedback, guiding learning through correct and incorrect reasoning responses (A92, A95–A97, A101–A102), and suggesting further study resources(A 95). When combined with analytics (A92, A96, A101, A102), these systems could yield nuanced insights for educators and learners on performance, identifying specific areas for improvement. Six studies provided Kirkpatrick level 1 (A92, A98, A101) and/or 2b evidence supporting the use of these systems (A92, A95-A96, A101–A102).

The creation of VPSs and ITSs cases, a labor-intensive process, typically involved content experts. However, 2 studies (A101, A 104) demonstrated innovative use of NLP to automate the creation of extensive case libraries (413 and 2525 cases, respectively) from actual hospital records and the New England Journal of Medicine Clinical Case Competition series. Such expansive libraries enhanced the variety of cases available to learners, addressing common challenges in teaching and assessing clinical reasoning, like case variability and the unpredictable nature of clinical encounters in real-world settings.

Further exploration into AI's utility is found in three studies that apply ML to analyze clinical reasoning in written notes (A93–A94, A99). These studies assessed different forms of clinical documentation, including post-encounter OCSE notes (A93) and diagnostic justification essays (A94) by medical students, as well as residents' admission notes from electronic health records using a standardized rubric (IDEA tool) for clinical notes (A99). The psychometric properties and inter-rater reliability of the ML algorithms were reported as acceptable. One study achieved a kappa statistic of 0.8, meeting the reliability standard for high-stakes assessment (A93), whereas others met the standard for lower-stakes assessments. The collective findings of these studies substantiate the concept of utilizing ML to streamline assessment of clinical reasoning, thereby enhancing efficiency of scoring and significantly enriching formative feedback through assessment of clinical notes–a largely underutilized source in clinical reasoning assessment.

The final two studies in this area highlighted AI's expanding role. The first study demonstrated the use of neural networks to recognize patterns of successful student problem-solving (A100), further highlighting AIs potential in clinical reasoning assessment. The second study evaluated the effectiveness of three different LLMs at providing clinical decision support for junior doctors (A103). The authors noted that more advanced LLMs trained on medical databases, rigorously scrutinized by medical experts, could result in valuable educational tools in the future.

AI use for automation of case and procedure logs

Four of the publications detailed efforts to utilize AI/ML to automate documentation of trainee clinical and procedural experiences. In one (A107), an NLP algorithm–based system identified clinical concepts in trainee clinical notes and mapped these onto expert–defined core clinical problems, with a precision of 92.3%. A similar system based on a commercial NLP product tracked case experiences for neurology residents, with a tripling of case experiences logged in a one–month timeframe (A105). A third system (termed 'Trove') (A106) used an NLP algorithm to tally classes of diagnoses based on resident–created imaging impressions and to populate a dashboard for residents with accuracies ranging from 93.2 to 97.0%. Finally, a system utilizing a deep neural network with a reinforcement learning model provided suggestions for data residents needed to manually log their surgical case experiences based on EHR data and surgical case schedules (A108).

Knowledge of AI

Fifty-one of the included articles (26.7%) explored knowledge, perceptions, attitudes, hopes, and concerns related to AI in medical education.

Three studies carried out objective assessments of AI knowledge. In Spain, a study of 281 medical students used true/false questions to assess general AI knowledge, with a 65% correct response rate (A117). In Saudi Arabia, another study of 476 medical students assessed knowledge of DL, and found that 54% failed to correctly answer any of five true/false questions (A118). Across Australia, New Zealand and the United States, a third study of 245 medical students evaluated ML knowledge through MCQs, resulting in an average score of 49.7% (A120).

Self-reported knowledge of AI varied. In Western Australia, 85% of medical students claimed basic understanding of AI (A151), and 91% of pediatric ophthalmologists reported familiarity with basic AI concepts (A154). In contrast, only 31% of German medical students reported basic knowledge of AI technologies (A145), and focus group discussion showed pediatric radiology trainees at a large academic hospital struggled with AI terminology and literature appraisal (A156). A South Korea survey indicated that medical students' primary information sources were news publications and TV (82%), social media (41%), lectures (32%), or friends or family (29%), with academic publications and books at only 22% (A127). Similarly, in a US

study, 72% of medical students and 59% of faculty learned about AI from media sources (A159).

AI-related course availability was inconsistent. In Germany, 71.8% of medical schools offered AI courses, mostly as electives or extra-curricular activities (A142). However, in a sample of Canadian medical students, 85% reported no formal AI education as part of the medical degree curriculum (A146).

Attitudes towards the application of AI in clinical medicine

The studies reviewed revealed mixed attitudes toward AI in clinical medicine. Most medical students and doctors acknowledged AI's potential benefits, such as enhancing clinical judgment, research, and auditing skills, and streamlining administrative processes (A111, A117, A141). However, concerns included the ethical implications, weakened physician-patient relationships, and AI's limitations in unexpected scenarios (A114, A119, A154, A140).

Two studies investigated medical students' intention to use AI. One used the Unified Theory of User Acceptance of Technology (UTAUT) to examine the intention to use AI for diagnosis support, finding social influence as the only key factor (A153). In the other study, intention to use AI was only observed when students had a strong belief in the role of AI in the future of the medical profession (A157).

Attitudes toward the application of AI in medical education

Two studies assessed attitudes on the use of AI in medical education. One study assessed Lebanese medical students' views about using AI for assessment (A131). While 58% believed AI would be more objective than a human examiner, only 26.5% preferred AI assessments. Most (71%) preferred a combination of AI and humans. Another study in a Caribbean medical school found that 33% of faculty used ChatGPT, mainly for generating MCQs. The main concerns about its use were the potential for incorporation of misinformation into teaching materials and plagiarism.

Concerns and hopes for AI

Concerns about AI included its potential to reduce the demand for doctors and influence career choices, especially in radiology (A113, A115, A124, A135, A137,A147, A150). Insufficient AI training in medical curricula was another worry (A128, A143, A151). Three studies assessed aspects of AI which should be included in medical curricula, including clinical applications, algorithm development, algorithm appraisal, coding, and basics of AI, statistics, ethics, and privacy (A133, A156, A158).

Performance of LLMs on medical exams

Thirty-two studies assessed various LLMs' abilities to perform on medical exams or on test questions drawn from large question banks (A160-A191). The LLMs tested included Microsoft's Bing Chat, ChatGPT, GPT-3, GPT-3.5, and GPT-4, Google's Bard, and several other pre-trained language models. They were tested on multiple choice (A160-A185, A188-A190), short answer and essay questions (A186-A187, A191), though questions with images and diagrams were typically excluded due to LLMs' current limitations processing visuals. A few studies asked the LLM to justify its answers (A164, A168, A175-A176), which were then assessed for accuracy. LLM performance was compared to varying levels of human performance (i.e. novice to expert), as well as 'gold standard' answers (A160-A191). LLMs scored at or above passing thresholds on a number of local, university-derived, basic and clinical science exams (A178, A183, A184), as well as on national licensure exams (A160, A169, A180, A182, A187, A189) and specialty licensure exams (A169, A175, A181). Despite these successes, LLMs performed only modestly or poorly on other exams (A168, A170, A173-A174, A176, A190). Newer generation LLMs generally outperformed older generation models, though this was not universally true. LLMs typically performed better on general medical exams, compared to specialty or subspecialty exams, though this was also not universal. LLM performance was tested on exams from Australia (A177), Canada (A162, A174), China (A173, A190), France (A160, A170, A181), Germany (A165, A182), India (A160, A163, A183-A184, A187), Italy (A160), Japan (A189), Korea (A171, A180), Netherlands (A179), Saudi Arabia (A178), Spain (A160), Switzerland (A185), Taiwan (A173), Turkey (A166), the UK (A160, A167, A191), and the US (A160-A161, A164, A168-A169, A172-A173, A175-A176, A186, A188). In most cases, LLMs performed better on English language questions, or when questions in other languages were first translated into English. Their performance in other languages was highly variable (e.g. 22% accuracy in French vs. 73% in Italian (A160, A170). At this time, learners and educators should approach use of LLMs with caution due to variable accuracy (A188), though the notable improvements of newer LLMs may soon overcome this barrier. As capabilities of LLMs improve, educators must contend with concerns about the integrity of formal assessments, as well as their reliability and validity. Assessments may require redesign, and a paradigm shift towards open book tests that emphasize learners' abilities to integrate information, rather than recall facts, may be warranted (A186).

Overview of perspectives publications included in the review

Eighty-seven perspective publications met inclusion criteria for this review, labeled P1–P87 in Supplementary Table 3. These perspectives, encompassing commentaries, editorials, and letters to the editor, offered insights on the use of AI in medical education, presenting opinions, recommendations, or examples of potential use cases of AI. We categorized these perspectives into five groups according to their primary content—general arguments for the importance of incorporating AI in medical education ($n = 42$); suggested educational applications or developments ($n = 18$); frameworks for curricular development ($n = 8$); recommendations for curricular content, AI competencies, or resources to support use of AI ($n = 15$); and ethics considerations ($n = 4$). Some publications had secondary content in an additional category (see Supplementary Table 3 and Supplementary Appendix 2 for additional details).

Importance of integrating AI into medical education

Many of the perspectives argued that incorporating AI in medical education is essential to prepare future physicians to apply AI to patient care (P50, P52–53, P60, P65–67, P70, P79). AI is already influencing diagnosis and treatment decisions in health care, and medical students and residents need to be ready to use AI tools effectively in the health care system. Additionally, several perspectives advocated for the use of AI to enhance teaching and learning (P55–56, P58–59, P61, P64, P69, P71, P84–86) or assessment (P48, P81). Two perspectives urged medical educators to incorporate teaching about AI's influence and ethical and legal considerations (P57, P87). Some of the perspectives raise cautions about AI's potential to compromise or limit clinical learning (P46, P49, P54, P63, P72–73, P82). The remaining perspectives contained general statements about the potential use of AI in medical education (P47, P49, P51, P54, P62, P68, P74–78, P80, P82–83).

Specific suggestions for educational applications and ideas for development

The 18 perspectives with specific educational applications included the following uses of AI tools in medical education:

- **Intelligent Tutoring Systems** – AI-powered platforms providing personalized learning experiences to improve the teaching of decision-making skills (P5)
- **AI-assisted learner assessment** – Latent Semantic Analysis to grade students' clinical case summaries and provide feedback (P11); NLP to score clinical skills exams (P15)
- **ChatBot** (e.g. ChatGPT) – AI processing and understanding of medical literature to teach clinical management (P10), assist with USMLE (P12) and other exam (P17) preparation, promote critical thinking, creativity, and patient communication (P8, P14, P16)
- **Personalized Learning Platforms** – AI-driven platforms creating personalized learning paths for students (P1) and providing personalized feedback (P8)
- **Robot-Assisted Surgery Simulations** (e.g. Virtual Reality) – AI-powered simulators for surgical training and evaluation (P13)
- **Enhanced anatomy education –** enhancing teaching, learning, and assessment in anatomy education for deeper learning and long-term retention (P2)
- **AI tools to prepare applications –** medical school applications (with a call for guidelines to ensure fairness in admissions) (P6) and personal statements for residency applications (P18)
- **AI-generated art** – to enhance visual storytelling for patient encounters (P7)
- **Machine learning and intra-operative video analysis to improve patient care** – teaching competency-based patient assessments (P9)

Frameworks for AI in medical education

Several perspectives provided frameworks to guide the incorporation of AI into medical education. An AI literacy framework for medical school described aims, opportunities, and impact of AI for selection into medical school and training programs, learning, assessment, and research (P19). A framework for radiology education described the role of AI to promote personalized learning and decision support tools (P20). A framework that emphasized digital skills provided a structure for teaching data protection, legal aspects of digital tools, digital communication, data literacy and analytics (P21). A framework for integrating AI in UME used a case-based approach with a detailed delineation of what AI content could be taught through cases (P22). A framework for early and progressive exposure to AI for medical students provided an analytics hierarchy to scaffold learning to prepare students to adapt to evolving clinical settings (P23). A framework for utilizing the rapidly expanding medical knowledge base also incorporated content about how AI relates to health care economics, regulations, and patient care (P24). An embedded AI ethics education framework provided a curricular structure for teaching the ethical dimensions of medical AI in the context of existing bioethics curricula. (P25). Finally, a curriculum framework incorporated ML to build clinicians' skills in research, applied statistics, and clinical care (P26).

Recommendations for AI curriculum content or competency, tools and resources

Fifteen perspectives offered explicit recommendations for curricular content about AI, AI competencies, and specific tools and resources that can support curriculum development for teaching about AI in medical education. Recommended content included what physicians will need to know to meet patients' needs as AI changes the practice of medicine (P31–32), data science and concepts underlying AI (P33, P35), using AI to improve surgical outcomes (P41), surgical training for ophthalmologists (P40), AI model performance, financial implications, and generation of reports in radiology (P38–39, P37), data science to support evidence-based medicine (P36), using AI for diagnostics and prognosis in hematology (P28), concepts underlying machine learning (P34), and ethical uses of AI in medicine (P27). One publication proposed a curriculum for a new specialty, Doctor in Medical Data Sciences (P29). One publication explained six domains of competency for effective use of AI tools in primary care, in the context of the Quintuple Aim (P30).

Ethics considerations for AI education

Fourteen perspectives addressed ethics considerations and reflected a multifaceted view of the ethical landscape AI introduces to healthcare. Many of these publications expressed caution about the limitations of AI applications in medical education. These perspectives also stressed the importance of educational approaches that equip healthcare professionals with the necessary skills to navigate AI challenges, emphasizing the irreplaceable value of human judgment and ethical reasoning in the age of AI.

Recommended focus for ethics education

Many of the perspectives papers raised concerns about the ethics of using AI in healthcare and medical education, and discussed the responsibility of physicians to grapple with the ethical challenges that can compromise education and health care. These papers emphasize the importance of

preparing current and future physicians to grapple with these issues, while acknowledging that we have neither fully grasped the ethical implications of AI nor learned how to adequately respond to the ethical challenges. When considered collectively, these papers contribute important insights about the topics that must be addressed in medical education. Referring to these papers can provide clear guidance to medical educators about ethical issues to include in curricula. The perspectives that addressed ethics related to AI recommended ethics education on the following topics, which are summarized in Figure 3:

- **Algorithmic Bias and Equity:** Future clinicians must understand biases in health AI, especially those caused by training on non-representative data sets, which might exacerbate healthcare disparities (P42). Through an AI curriculum, physicians can reflect critically on their assumptions and the biases in both AI and clinical practice. This could help them better interpret AI outputs and recognize the limitations of current medical knowledge (P22).
- **Resource Allocation**: AI in healthcare brings to the fore issues of just resource allocation and equity, challenging students to consider how AI can be designed and used to promote justice in healthcare (P42).
- **Safety and Quality Assurance:** AI necessitates agreement among healthcare professionals to use AI tools responsibly and in alignment with professional obligations, such as practicing medicine safely and effectively (P25). With AI's integration into clinical practice, students must understand regulatory efforts, gaps in AI safety, and the ethical implications of using such technologies (P42).
- **Human Interaction and Compassionate Care:** AI-based ethics training might be less likely to instill compassion in students compared to traditional methods. One publication argued that compassion can indeed be taught and that positive role models in medical education can deepen students' understanding of compassionate practice (P44).
- **Data Privacy and Security**: The use of AI in medicine necessitates a strong understanding of data privacy, the ethical handling of patient data, and the legal dimensions associated with AI (P42, P43). The large number of available data sets and increased power of AI capabilities make data de-anonymization (or reidentification) highly possible through cross-referencing with other data sets (P43).
- **Automation Bias and Skill Preservation**: Clinicians must be wary of automation bias, where undue reliance on AI could lead to errors, highlighting the need for balanced trust and skepticism (P25). Ethics education must ensure AI complements, not replaces, the clinician's role in compassionate care (P25). Training should focus on preserving clinical skills and human interactions that AI cannot replicate (P42), leveraging 'hybrid intelligence' systems to strengthen therapeutic patient-provider relationship and empathy (P44).
- **Transparency and Informed Consent**: The deployment of health AI raises critical questions about informed consent and the physician's duty to inform patients about AI's involvement in their care decisions,

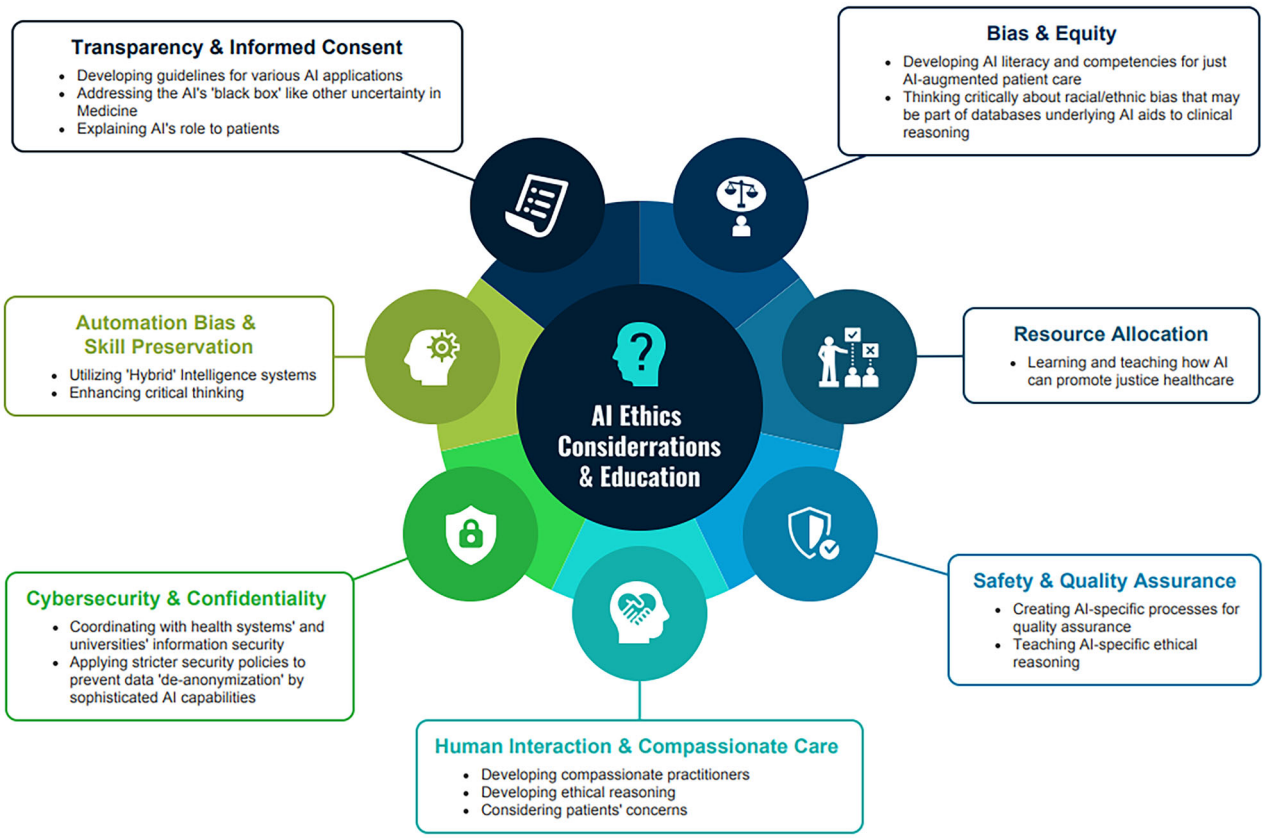

Figure 3. Ethical considerations for AI in medical education.

highlighting the need for medical students to engage in these discussions (P42). Proficiency in medical AI is intertwined with ethical competency, where ethical considerations significantly influence clinical decisions and vice versa. As future practitioners, students need to learn to identify AI's potential consequences and how to discuss them with patients in light of ethical principles (P25). In medical education, students must be made aware of data collection, provide consent, and understand the extent and implications of their consent (P43).

Proposed educational approaches for ethics education

One recommended approach to teaching ethics about AI is the incorporation of real-life case studies in medical curricula to elucidate the multifaceted ethical issues presented by health AI, such as informed consent, bias, and transparency (P42). This case-based methodology not only enriches the theoretical understanding but also hones practical decision-making skills. An Embedded AI Ethics Education Framework, which is an incremental approach that integrates AI ethics within existing medical ethics courses, addresses the potential harms of technology misuse and equips students to conduct risk-benefit analyses fundamental to patient care (P25). A third approach uses AI scenarios currently employed in clinical practice as examples for case-based learning to prepare students to critically evaluate the technical aspects of AI alongside ethical implications (P22). This approach fosters critical reflection and encourages interdisciplinary dialogue essential for the application of AI in patient-centered care. Another approach emphasizes the limitations of AI through human-led ethics teaching, much like ethics teaching that has become commonplace in medical school curricula. This approach maintains an emphasis on human interaction and the cultivation of moral reasoning over reliance on technology, while remaining open to the potential for 'hybrid' intelligence systems, in which humans and machines learn from each other, resulting in superior performance in ethical decision-making. Collectively, these educational strategies underscore the necessity of equipping healthcare professionals with both technical skill and ethical discernment to navigate the complexities introduced by AI in healthcare and health professions education.

Discussion

This review embarked on a journey to map the explosion of interest in AI across the medical education landscape through the lens of published works. The detailed tables of articles and perspectives papers, summary infographics and associated narrative clarify the current state of the evidence. The first published study on AI in medical education dated back three decades to 1992, though AI has existed much longer than this. A steady flow of publications has been seen since, a testament to the enduring fascination and steady evolution of AI in medical education. The recent surge in publications may at first appear to be a growth spurt associated with maturation in the field, though our analysis suggests it is more likely a contemporaneous response to the public release of ChatGPT.

The recent advancements in AI technology have led to rapid, widespread and in many ways limitless distribution of accessible, cost-effective AI tools globally. This technological democratization is mirrored in the diverse origin of publications with almost a third from regions outside of the traditional North American or European academic hubs, more than found in most recently published BEME reviews (Daniel et al. 2022; Doyle et al. 2023; Hamza et al. 2023; Li S et al. 2023). This review reflects a shift from centralized to distributed perspectives in AI research and development, offering a more inclusive and comprehensive understanding of AI's role in shaping medical education worldwide.

This scoping review, in charting the course of AI in medical education, reveals not just the peaks and troughs of evidence, but also an expansive change that has permeated every facet of the educational environment. When considering the impact at such a broad environmental level, our analysis, viewed through the conceptual lens of Oberg's theory of 'culture shock' (Oberg 1960; Zhou et al. 2008), aligns with the theory's four phases: Honeymoon, Frustration, Adaptation, and Acceptance. Each phase represents a distinct emotional, psychological, and behavioral response to AI integration manifesting variably, across different areas and individuals.

The initial 'Honeymoon' phase is characterized by a mixture of fascination and idealism, often reflected in perspective and innovation pieces. These works, often written by experts, provide valuable insights, useful resources and some directions or foundations to move the field forward. These sometimes portray an overtly optimistic view of AI's potential, lacking detailed exploration of challenges and ethical considerations.

In contrast, the 'Frustration' phase has been seen in evidence of attitudes of feeling overwhelmed, fear related to lack of understanding, and feeling lost and out of place. These concerns are not only valid but a necessary step in progressing towards acceptance of such a shock in the system. The ethical considerations and challenges related to specific applications of AI are key to consider. Balancing these aspects is crucial for sustainable implementation before the next phase.

The ambivalence observed in a large group of studies examining NLP, mostly ChatGPT, for medical exams [A160–191], exemplifies the non-linear and individual nature of cultural adaptation. This straddling of both Honeymoon and Frustration phases reflects a duality – awe of AI's capabilities juxtaposed with apprehension about its implications, particularly in reducing the need for traditional medical education methods. We draw attention to this group of publications not just as an interesting example, but to highlight that it is clear that without an innovative or different perspective, no further studies of this form appear justified.

As the initial frustrations and challenges of AI integration begin to crystallize, we transition into the 'Adaptation' phase. Educators begin to find practical and effective ways to incorporate AI tools. This phase is marked by the pragmatic 'use cases' of AI, showing a shift from initial awe and skepticism towards AI to a more balanced and functional integration of these technologies. However, discussions around these use cases in perspective pieces often lack

depth in building a longitudinal or iterative understanding, which is critical for effective integration.

Finally, in the 'Acceptance' phase, we witness a transformation in the educators' relationship with AI, evolving from mere adaptation to a state of genuine integration and comfort. This phase signifies a shift where AI is no longer viewed as an external, disruptive force but as an integral part of the educational blueprint. The subset of publications exhibits a deep understanding and skillful utilization of AI, seamlessly incorporating these technologies into their wide-ranging educational developments, characterized by a blend of creativity, critical thinking, and a forward-looking perspective that acknowledges the potential of AI while remaining cognizant of its limitations and ethical considerations. This mature phase of acceptance is not merely about recovery from the initial 'shock' of AI integration; it represents a proactive, thoughtful embrace of these technologies. These educators have become pioneers in a true sense, setting benchmarks for how AI can be harmoniously and innovatively integrated into medical education. Their experiences and practices offer valuable insights for the broader educational community, illustrating a roadmap from apprehension to acceptance, a journey that is as much about personal and professional growth as it is about technological adoption.

In the process of extracting and charting the data, we have identified critical gaps in how a study is reported – with a need for a 'compass' to navigate the roadmap in AI applications in medical education. Here, addressing gaps involves a shift from mere reporting of disparate items, such as educational contexts, stages, or AI technologies, to a comprehensive framework that supports dissemination, replication, and innovation through several elements. We propose the FACETS framework, presented in Figure 4, as a structured approach for future AI studies in medical education, focusing on key elements: form, use case, context, education form, technology, and SAMR framework for technology integration. Our review reveals that these elements are commonly present in the literature, but they are typically reported in isolation, not cohesively in a single publication. Bringing together these disparate elements in a more systematic and unified manner is valuable in explaining to both educators and education researchers the specific details of the AI being employed, their relevance to past practice, and how they can be integrated into the educational context. It is designed to not only guide authors in conceptualizing their works but also to help readers to grasp the full scope and impact of AI in medical education. This approach is essential for advancing understanding and innovation in the field, forming a core recommendation from the author team for future research.

This framework cannot be applied to perspective pieces or attitudes surveys, but the use of this framework by authors of works published regarding any AI use-case or broader innovation will be of significant benefit. The FACETS framework may aid in identifying converging areas of research and serve as a useful guide for more focused syntheses in the future.

As this review was rigorous, systematic, and wide ranging in true alignment with the scoping tradition, we do not believe further scoping reviews are indicated. However, given the velocity of output, more in-depth systematic reviews will soon be indicated in the areas of admissions, teaching, assessment, and clinical reasoning. Another area where a systematic review may be warranted was out of scope for this review, namely, AI use in medical education research. Given the huge untapped potential of big data in educational institutions, this is a key area for further work.

Limitations of this review include the limitations of the scoping tradition, which by definition, does not synthesise evidence (Levac et al. 2010). Further limitations include subjectivity in categorizing studies (e.g. some studies could be classified as both teaching and assessment), as well as issues with consistent terminology, which underscore the complexity of navigating a rapidly advancing field like AI in medical education. The review iterated the search on a number of occasions in the early phases, reflecting the vast and evolving scope of AI literature. In light of this, studies may have been missed, despite systematic and rigorous

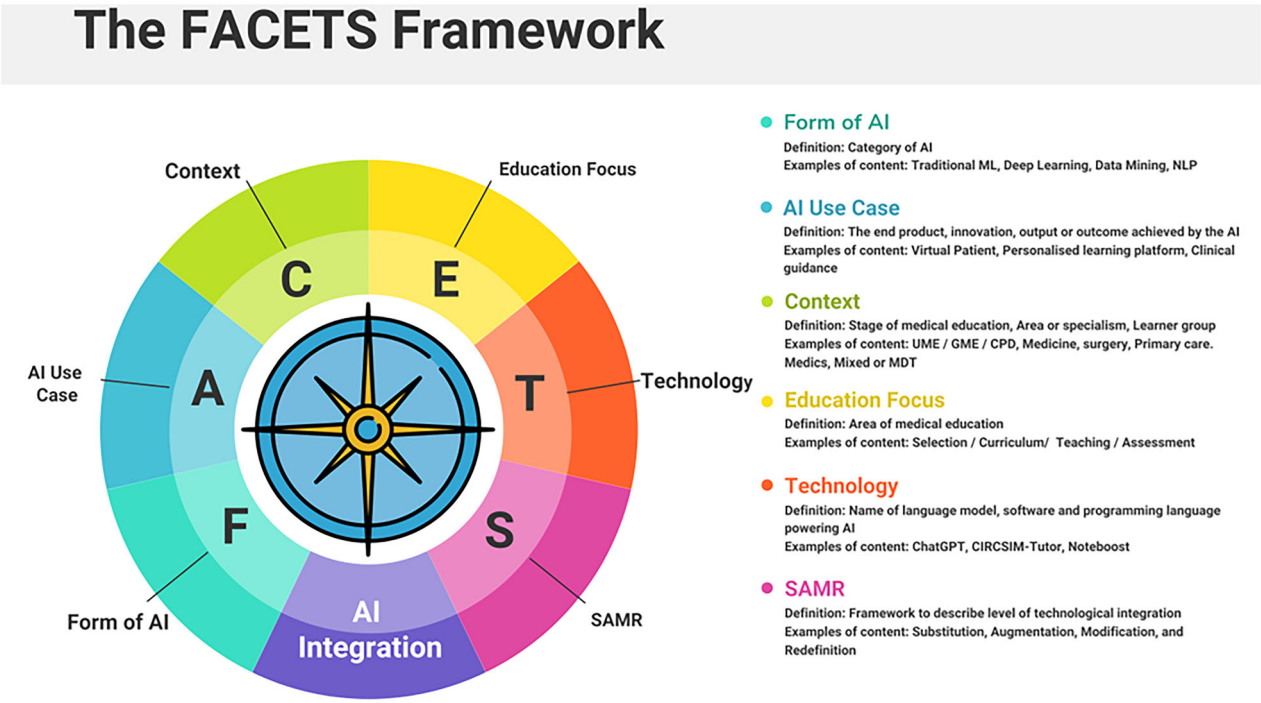

Figure 4. FACETS framework to report future AI innovations in medical education.

methods. The inclusion of perspective articles, while unconventional for BEME reviews, offers a diverse perspective on AI in medical education. However, this inclusion may be perceived as diluting the empirical strength of the review. Thus, a decision was made to present their synthesis transparently as a separate parallel data set within the results, as there was much value in clarifying what these articles offered (and did not offer). We did not perform the sixth stage of the Arksey and O'Malley scoping review framework. While we had AI and education experts on the author team who fulfilled this role, skipping the external expert consultation step could be viewed as a limitation.

Implications for practice and policy

This review underscores the transformative role of AI in medical education, necessitating an integration of AI into various aspects like curriculum development, teaching methodologies, and learner assessment to maintain alignment with technological advancements. It calls for updating competency frameworks and establishing standardized guidelines for AI usage in education. The need for collaborative efforts between educators, policymakers, and AI developers is highlighted to ensure effective integration and address equity and accessibility issues, ensuring AI tools are available to all without bias. Continuous investment in resources and time is essential to transition from initial fascination with AI to its profound adoption in educational settings. The review also stresses managing risks associated with AI, including ethical concerns and potential misuse. Crucially, while leveraging AI for new insights and efficiencies, the irreplaceable human element in healthcare must be preserved. Ethical education is vital, ensuring AI enhances rather than replaces the healthcare provider's role, upholding patient-provider relationships, empathy, and respect for patient autonomy. This approach prevents the dehumanization of care and ensures AI's use bolsters rather than undermines the core human skills and values central to the medical profession.

Implications for future research

The review highlights several key areas for future research and educational development. It underscores the need for comprehensive studies on the long-term impacts of AI, both on the overall landscape of medical education and specific learning outcomes. Additionally, there is a call for research addressing potential risks associated with AI, such as automation bias, over- and under-skilling, exacerbating disparities, and the preservation of essential clinical skills. It also recommends employing the FACETS framework for future AI studies in medical education. This comprehensive approach to research and development will ensure that AI is integrated into medical education in a manner that is ethically sound, educationally effective, and cognizant of the potential risks and challenges.

Conclusions

The landscape of AI in medical education, as charted in this review, spans a wide array of stages, specialties, purposes, and use cases, primarily reflecting early adaptation phases–only a few describe more in-depth employment for longitudinal or deep change. The proposed FACETS framework is a key outcome, offering a structured approach for future research and practice. While no immediate areas for systematic review are identified, the dynamic nature of AI in education necessitates ongoing vigilance and adaptability.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Notes on contributors

Morris Gordon, MBChB, PHD, MMed, is Co-chair of the BEME Council, Coordinating Editor of the Cochrane Gut Group, Professor of Evidence Synthesis and Systematic Review and a Consultant Pediatrician in the North West of the UK.

Michelle Daniel, MD, MHPE, is Co-chair of the BEME Council, Associate Editor for Medical Teacher, Vice Dean for Medical Education and Clinical Professor of Emergency Medicine at the University of California, San Diego School of Medicine, La Jolla, California, USA.

Aderonke Ajiboye, BDS, MPH, is a Researcher at the School of Medicine and Dentistry, University of Central Lancashire.

Hussein Uraiby, MBChB, MMedEd, FHEA, is a Specialty Trainee in Histopathology at the University Hospitals of Leicester NHS Trust, Leicester, UK.

Nicole Y. Xu is a medical student at the University of California, San Diego School of Medicine, La Jolla, California, USA.

Rangana Bartlett is a recent graduate of the University of California, San Diego and current research assistant at the Icahn School of Medicine at Mount Sinai, New York, New York, USA.

Janice Hanson, PhD, EdS, MH, is Professor of Medicine, Co-Director of the Medical Education Research Unit, and Director of Education Scholarship Development at Washington University in Saint Louis, Missouri, USA.

Mary Haas, MD, MHPE, is Assistant Residency Program Director and Assistant Professor, Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA.

Maxwell Spadafore, MD, has expertise in AI and informatics and is a Chief Resident, Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA.

Ciaran Grafton-Clarke, MBChB, is an NIHR Academic Clinical Fellow in Cardiology, University of East Anglia, Norwich, UK.

Rayhan Yousef Gasiea, MBBS, is a junior clinical fellow in medical education at Blackpool Hospitals NHS Foundation Trust.

Collin Michie, FRCPCH, FRSPH, FLS, RNutr, is Associate Dean for Research and Knowledge Exchange at the School of Medicine and Dentistry, University of Central Lancashire, Preston, UK.

Janet Corral, EdD is Professor of Medicine and Associate Dean of Undergraduate Medical Education at the University of Nevada School of Medicine, Reno, Nevada, USA.

Brian Kwan, MD has advanced training in informatics, and is an Assistant Clinical Professor and Medical Director for Education Informatics at UC San Diego School of Medicine, USA.

Diana Dolmans, PhD, is a Professor and Chair of the Department of Educational Development & Research, staff member of the School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, NL.

Satid Thammasitboon, MD, MHPE, is Director of the Center for Research, Innovation and Scholarship in Health Professions Education, Texas Children's Hospital and Baylor College of Medicine, Houston, Texas, USA, and member of the BEME Council.

ORCID

Morris Gordon http://orcid.org/0000-0002-1216-5158
Michelle Daniel http://orcid.org/0000-0001-8961-7119
Nicole Y. Xu http://orcid.org/0000-0001-9353-8819
Brian Kwan http://orcid.org/0000-0002-5835-0631

References

Abbott KL, George BC, Sandhu G, Harbaugh CM, Gauger PG, Ötleş E, Matusko N, Vu JV. 2021. Natural language processing to estimate clinical competency committee ratings. J Surg Educ. 78(6):2046–2051. doi: 10.1016/j.jsurg.2021.06.013.

Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, Aziz S, Damseh R, Alrazak SA, Sheikh J. 2023. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ. 9(1):e48291. doi: 10.2196/48291.

Abdellatif H, Al Mushaiqri M, Albalushi H, Al-Zaabi AA, Roychoudhury S, Das S. 2022. Teaching, learning and assessing anatomy with artificial intelligence: the road to a better future. Int J Environ Res Public Health. 19(21):14209. doi: 10.3390/ijerph192114209.

Abid S, Awan B, Ismail T, Sarwar N, Sarwar G, Tariq M, Naz S, Ahmed A, Farhan M, Uzair M, et al. 2019. Artificial intelligence: medical students' attitude in district Peshawar Pakistan. PJPH. 9(1):19–21. doi: 10.32413/pjph.v9i1.295.

Afzal S, Dhamecha TI, Gagnon P, Nayak A, Shah A, Carlstedt-Duke J, Pathak S, Mondal S, Gugnani A, Zary N, et al. 2020. AI medical school tutor: modelling and implementation. AIME 2020. Proceedings of the 18th International Conference on Artificial Intelligence in Medicine; Aug 25–28; Minnesota. Springer International Publishing; p. 133–145. doi: 10.1007/978-3-030-59137-3_13.

Agarwal M, Sharma P, Goswami A. 2023. Analysing the applicability of ChatGPT, Bard, and Bing to generate reasoning-based multiple-choice questions in medical physiology. Cureus. 15(6):e40977. doi: 10.7759/cureus.40977.

Agha-Mir-Salim L, Mosch L, Klopfenstein SA, Wunderlich MM, Frey N, Poncette AS, Balzer F. 2022. Artificial intelligence competencies in postgraduate medical training in Germany. Challenges of trustable AI and added-value on health [e-book]. Amsterdam: IOS Press; p. 805–806. doi: 10.3233/shti220589.

Ahmed Z, Bhinder KK, Tariq A, Tahir MJ, Mehmood Q, Tabassum MS, Malik M, Aslam S, Asghar MS, Yousaf Z. 2022. Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in Pakistan: a cross-sectional online survey. Ann Med Surg. 76:103493. doi: 10.1016/j.amsu.2022.103493.

Ahmer H, Altaf SB, Khan HM, Bhatti IA, Ahmad S, Shahzad S, Naseem S. 2023. Knowledge and perception of medical students towards the use of artificial intelligence in healthcare. J Pak Med Assoc. 73(2):448–451. doi: 10.47391/jpma.5717.

Ahuja AS, Polascik BW, Doddapaneni D, Byrnes ES, Sridhar J. 2023. The digital metaverse: applications in artificial intelligence, medical education, and integrative health. Integr Med Res. 12(1):100917. doi: 10.1016/j.imr.2022.100917.

Alamer A. 2023. Medical students' perspectives on artificial intelligence in radiology: the current understanding and impact on radiology as a future specialty choice. Curr Med Imaging. 19(8):921–930. doi: 10.2174/1573405618666220907111422.

Aldeman NL, de Sá Urtiga Aita KM, Machado VP, da Mata Sousa LC, Coelho AG, da Silva AS, da Silva Mendes AP, de Oliveira Neres FJ, do Monte SJ. 2021. Smartpathk: a platform for teaching glomerulopathies using machine learning. BMC Med Educ. 21(1):248. doi: 10.1186/s12909-021-02680-1.

Alfertshofer M, Hoch CC, Funk PF, Hollmann K, Wollenberg B, Knoedler S, Knoedler L. 2023. Sailing the seven seas: a multinational comparison of ChatGPT's performance on medical licensing examinations. Ann Biomed Eng. 8:1–4. doi: 10.1007/s10439-023-03338-3.

Ali R, Tang OY, Connolly ID, Sullivan PL, Shin JH, Fridley JS, Asaad WF, Cielo D, Oyelese AA, Doberstein CE, et al. 2023. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery. 93(6):1353–1365. doi: 10.1227/neu.0000000000002632.

Alonso-Silverio GA, Pérez-Escamirosa F, Bruno-Sanchez R, Ortiz-Simon JL, Muñoz-Guerrero R, Minor-Martinez A, Alarcón-Paredes A. 2018. Development of a laparoscopic box trainer based on open source hardware and artificial intelligence for objective assessment of surgical psychomotor skills. Surg Innov. 25(4):380–388. doi: 10.1177/1553350618777045.

Al Saad MM, Shehadeh A, Alanazi S, Alenezi M, Abu Alez A, Eid H, Alfaouri MS, Aldawsari S, Alenezi R. 2022. Medical students' knowledge and attitude towards artificial intelligence: an online survey. TOPHJ. 15(1):1–8. doi: 10.2174/18749445-v15-e2203290.

Alverson DC, Saiki SM, Jr, Caudell TP, Goldsmith T, Stevens S, Saland L, Colleran K, Brandt J, Danielson L, Cerilli L. 2006. Reification of abstract concepts to improve comprehension using interactive virtual environments and a knowledge-based design: a renal physiology model. Stud Health Technol Inform. 119:13–18.

AlZaabi A, AlMaskari S, AalAbdulsalam A. 2023. Are physicians and medical students ready for artificial intelligence applications in healthcare? Digit Health. 9:20552076231152167. doi: 10.1177/20552076231152167.

Antaki F, Touma S, Milad D, El-Khoury J, Duval R. 2023. Evaluating the performance of chatgpt in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci. 3(4):100324. doi: 10.1016/j.xops.2023.100324.

Arksey H, O'Malley L. 2005. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 8(1):19–32. doi: 10.1080/1364557032000119616.

Armitage RC. 2023. ChatGPT: the threats to medical education. Postgrad Med J. 99(1176):1130–1131. doi: 10.1093/postmj/qgad046.

Arora VM. 2018. Harnessing the power of big data to improve graduate medical education: big idea or bust? Acad Med. 93(6):833–834. doi: 10.1097/acm.0000000000002209.

Asghar A, Patra A, Ravi KS. 2022. The potential scope of a humanoid robot in anatomy education: a review of a unique proposal. Surg Radiol Anat. 44(10):1309–1317. doi: 10.1007/s00276-022-03020-8.

Aubignat M, Diab E. 2023. Artificial intelligence and ChatGPT between worst enemy and best friend: the two faces of a revolution and its impact on science and medical schools. Rev Neurol. 179(6):520–522. doi: 10.1016/j.neurol.2023.03.004.

Auloge P, Garnon J, Robinson JM, Dbouk S, Sibilia J, Braun M, Vanpee D, Koch G, Cazzato RL, Gangi A. 2020. Interventional radiology and artificial intelligence in radiology: is it time to enhance the vision of our medical students? Insights Imaging. 11(1):127. doi: 10.1186/s13244-020-00942-y.

Ayub I, Hamann D, Hamann CR, Davis MJ. 2023. Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: a qualitative analysis. Cureus. 15(8):e43717. doi: 10.7759/cureus.43717.

Bakshi SK, Lin SR, Ting DS, Chiang MF, Chodosh J. 2021. The era of artificial intelligence and virtual reality: transforming surgical education in ophthalmology. Br J Ophthalmol. 105(10):1325–1328. doi: 10.1136/bjophthalmol-2020-316845.

Baloul MS, Yeh VJ, Mukhtar F, Ramachandran D, Traynor MD, Jr Shaikh N, Rivera M, Farley DR. 2022. Video commentary & machine learning: tell me what you see, I tell you who you are. J Surg Educ. 79(6):e263-72–e272. doi: 10.1016/j.jsurg.2020.09.022.

Banerjee M, Chiew D, Patel KT, Johns I, Chappell D, Linton N, Cole GD, Francis DP, Szram J, Ross J, et al. 2021. The impact of artificial intelligence on clinical education: perceptions of postgraduate trainee doctors in London (UK) and recommendations for trainers. BMC Med Educ. 21(1):1–0. doi: 10.1186/s12909-021-02870-x.

Bansal M, Jindal A. 2022. Artificial intelligence in healthcare: should it be included in the medical curriculum? A students' perspective. Natl Med J India. 35(1):56–58. doi: 10.25259/nmji_208_20.

Barreiro-Ares A, Morales-Santiago A, Sendra-Portero F, Souto-Bayarri M. 2023. Impact of the rise of artificial intelligence in radiology: what do students think? Int J Environ Res Public Health. 20(2):1589. doi: 10.3390/ijerph20021589.

Bhanvadia S, Saseendrakumar BR, Guo J, Daniel M, Lander L, Baxter SL. 2022. Evaluation of bias in medical student clinical clerkship evaluations using natural language processing. Acad Med. 97(11S): S154–S154. doi: 10.1097/ACM.0000000000004807.

Bhattacharya P, Van Stavern R, Madhavan R. 2010. Automated data mining: an innovative and efficient web-based approach to

maintaining resident case logs. J Grad Med Educ. 2(4):566–570. doi: 10.4300/jgme-d-10-00025.1.

Bin Dahmash A, Alabdulkareem M, Alfutais A, Kamel AM, Alkholaiwi F, Alshehri S, Al Zahrani Y, Almoaiqel M. 2020. Artificial intelligence in radiology: does it impact medical students preference for radiology as their future career? BJR Open. 2(1):20200037. doi: 10.1259/bjro.20200037.

Bisdas S, Topriceanu CC, Zakrzewska Z, Irimia AV, Shakallis L, Subhash J, Casapu MM, Leon-Rojas J, Pinto dos Santos D, Andrews DM, et al. 2021. Artificial intelligence in medicine: a multinational multi-center survey on the medical and dental students' perception. Front Public Health. 9:795284. doi: 10.3389/fpubh.2021.795284.

Blacketer C, Parnis R, B Franke K, Wagner M, Wang D, Tan Y, Oakden-Rayner L, Gallagher S, Perry SW, Licinio J, et al. 2021. Medical student knowledge and critical appraisal of machine learning: a multicentre international cross-sectional study. Intern Med J. 51(9):1539–1542. doi: 10.1111/imj.15479.

Blease C, Kharko A, Bernstein M, Bradley C, Houston M, Walsh I, Mandl KD. 2023. Computerization of the work of general practitioners: mixed methods survey of final-year medical students in Ireland. JMIR Med Educ. 9(1):e42639. doi: 10.2196/42639.

Blease C, Kharko A, Bernstein M, Bradley C, Houston M, Walsh I, Hägglund M, DesRoches C, Mandl KD. 2022. Machine learning in medical education: a survey of the experiences and opinions of medical students in Ireland. BMJ Health Care Inform. 29(1):e100480. doi: 10.1136/bmjhci-2021-100480.

Boillat T, Nawaz FA, Rivas H. 2022. Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ. 8(2):e34973. doi: 10.2196/34973.

Bond WF, Zhou J, Bhat S, Park YS, Ebert-Allen RA, Ruger RL, Yudkowsky R. 2023. Automated patient note grading: examining scoring reliability and feasibility. Acad Med. 98(11S):S90–S97. doi: 10.1097/acm.0000000000005357.

Booth GJ, Ross B, Cronin WA, McElrath A, Cyr KL, Hodgson JA, Sibley C, Ismawan JM, Zuehl A, Slotto JG, et al. 2023. Competency-based assessments: leveraging artificial intelligence to predict subcompetency content. Acad Med. 98(4):497–504. doi: 10.1097/acm.0000000000005115.

Borchert RJ, Hickman CR, Pepys J, Sadler TJ. 2023. Performance of ChatGPT on the situational judgement test - a professional dilemmas–based examination for doctors in the United Kingdom. JMIR Med Educ. 9(1):e48978. doi: 10.2196/48978.

Boscardin CK, Gin B, Golde PB, Hauer KE. 2023. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med. 99(1):22–27. doi: 10.1097/acm.0000000000005439.

Brandes GIG, D'Ippolito G, Azzolini AG, Meirelles G. 2020. Impact of artificial intelligence on the choice of radiology as a specialty by medical students from the city of São Paulo. Radiol Bras. 53(3):167–170. doi: 10.1590/0100-3984.2019.0101.

Brown JD, Kuchenbecker KJ. 2023. Effects of automated skill assessment on robotic surgery training. Int J Med Robot Comput Assist Surg. 19(2):e2492. doi: 10.1002/rcs.2492.

Buabbas AJ, Miskin B, Alnaqi AA, Ayed AK, Shehab AA, Syed-Abdul S, Uddin M. 2023. Investigating students' perceptions towards artificial intelligence in medical education. Healthcare. 11(9):1298. doi: 10.3390/healthcare11091298.

Burk-Rafel J, Reinstein I, Feng J, Kim MB, Miller LH, Cocks PM, Marin M, Aphinyanaphongs Y. 2021. Development and validation of a machine learning-based decision support tool for residency applicant screening and review. Acad Med. 96(11S):S54–S61. doi: 10.1097/acm.0000000000004317.

Busch F, Adams LC, Bressem KK. 2023. Biomedical ethical aspects towards the implementation of artificial intelligence in medical education. Med Sci Educ. 33(4):1007–1012. doi: 10.1007/s40670-023-01815-x.

Çalışkan SA, Demir K, Karaca O. 2022. Artificial intelligence in medical education curriculum: an e-Delphi study for competencies. PLoS One. 17(7):e0271872. doi: 10.1371/journal.pone.0271872.

Caparros GC, Portero FS. 2022. Medical students' perceptions of the impact of artificial intelligence in radiology. Radiología. 64(6):516–524. doi: 10.1016/j.rx.2021.03.006.

Carin L. 2020. On artificial intelligence and deep learning within medical education. Acad Med. 95(11S Association of American Medical Colleges Learn Serve Lead: proceedings of the 59th Annual Research in Medical Education Presentations):S10–S11. doi: 10.1097/acm.0000000000003630.

Chai SY, Hayat A, Flaherty GT. 2022. Integrating artificial intelligence into haematology training and practice: opportunities, threats and proposed solutions. Br J Haematol. 198(5):807–811. doi: 10.1111/bjh.18343.

Chen H, Gangaram V, Shih G. 2019. Developing a more responsive radiology resident dashboard. J Digit Imaging. 32(1):81–90. doi: 10.1007/s10278-018-0123-6.

Cheng C-T, Chen C-C, Fu C-Y, Chaou C-H, Wu Y-T, Hsu C-P, Chang C-C, Chung I-F, Hsieh C-H, Hsieh M-J, et al. 2020. Artificial intelligence-based education assists medical students' interpretation of hip fracture. Insights Imaging. 11(1):119. doi: 10.1186/s13244-020-00932-0.

Cheung BH, Lau GK, Wong GT, Lee EY, Kulkarni D, Seow CS, Wong R, Co MT. 2023. ChatGPT versus human in generating medical graduate exam multiple choice questions - a multinational prospective study (Hong Kong SAR, Singapore, Ireland, and the United Kingdom). PLoS One. 18(8):e0290691. doi: 10.1371/journal.pone.0290691.

Cho SI, Han B, Hur K, Mun JH. 2021. Perceptions and attitudes of medical students regarding artificial intelligence in dermatology. J Eur Acad Dermatol Venereol. 35(1):e72-3–e73. doi: 10.1111/jdv.16812.

Choi J, Lee Y, Kang GH, Jang YS, Kim W, Choi HY, Kim JG. 2022. Educational suitability of new channel-type video-laryngoscope with AI-based glottis guidance system for novices wearing personal-protective-equipment. Medicine. 101(9):e28890. doi: 10.1097/md.0000000000028890.

Cianciolo AT, LaVoie N, Parker J. 2021. Machine scoring of medical students' written clinical reasoning: initial validity evidence. Acad Med. 96(7):1026–1035. doi: 10.1097/acm.0000000000004010.

Civaner MM, Uncu Y, Bulut F, Chalil EG, Tatli A. 2022. Artificial intelligence in medical education: a cross-sectional needs assessment. BMC Med Educ. 22(1):772. doi: 10.1186/s12909-022-03852-3.

Clarke V, Braun V. 2013. Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. Psychol. 26(2):1–13.

Collado-Mesa F, Alvarez E, Arheart K. 2018. The role of artificial intelligence in diagnostic radiology: a survey at a single radiology residency training program. J Am Coll Radiol. 15(12):1753–1757. doi: 10.1016/j.jacr.2017.12.021.

Cooper A, Rodman A. 2023. AI and medical education-a 21st-century Pandora's box. N Engl J Med. 389(5):385–387. doi: 10.1056/nejmp2304993.

Coppin B. 2004. Artificial intelligence illuminated. Sudbury (MA): Jones & Bartlett Publishers.

Cross J, Robinson R, Devaraju S, Vaughans A, Hood R, Kayalackakom T, Honnavar P, Naik S, Sebastian R, Robinson RE. 2023. Transforming medical education: assessing the integration of ChatGPT into faculty workflows at a Caribbean medical school. Cureus. 15(7):e41399. doi: 10.7759/cureus.41399.

Cussat-Blanc S, Castets-Renard C, Monsarrat P. 2022. Doctors in medical data sciences: a new curriculum. Int J Environ Res Public Health. 20(1):675. doi: 10.3390/ijerph20010675.

Daniel M, Gordon M, Patricio M, Hider A, Pawlik C, Bhagdev R, Ahmad S, Alston S, Park S, Pawlikowska T, et al. 2021. An update on developments in medical education in response to the COVID-19 pandemic: a BEME scoping review: BEME guide no. 64. Med Teach. 43(3):253–271. doi: 10.1080/0142159X.2020.1864310.

Daniel M, Gottlieb M, Wooten D, Stojan J, Haas MRC, Bailey J, Evans S, Lee D, Goldberg C, Fernandez J, et al. 2022. Virtual interviewing for graduate medical education recruitment and selection: a BEME systematic review: BEME Guide No. 80. Med Teach. 44(12):1313–1331. doi: 10.1080/0142159X.2022.2130038.

Das D, Kumar N, Longjam LA, Sinha R, Roy AD, Mondal H, Gupta P. 2023. Assessing the capability of ChatGPT in answering first-and second-order knowledge questions on microbiology as per competency-based medical education curriculum. Cureus. 15(3):e36034. doi: 10.7759/cureus.36034.

Dave T, Athaluri SA, Singh S. 2023. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 6:1169595. doi: 10.3389/frai.2023.1169595.

Denny JC, Bastarache L, Sastre EA, Spickard A. III 2009. Tracking medical students' clinical experiences using natural language processing. J Biomed Inform. 42(5):781–789. doi: 10.1016/j.jbi.2009.02.004.

Denny JC, Spickard A, III, Speltz PJ, Porier R, Rosenstiel DE, Powers JS. 2015. Using natural language processing to provide personalized learning opportunities from trainee clinical notes. J Biomed Inform. 56:292–299. doi: 10.1016/j.jbi.2015.06.004.

Doumat G, Daher D, Ghanem NN, Khater B. 2022. Knowledge and attitudes of medical students in Lebanon toward artificial intelligence: a national survey study. Front Artif Intell. 5:1015418. doi: 10.3389/frai.2022.1015418.

Doyle AJ, Sullivan C, O'Toole M, Tjin A, Simiceva A, Collins N, Murphy P, Anderson MJ, Mulhall C, Condron C, et al. 2023. Training simulated participants for role portrayal and feedback practices in communication skills training: a BEME scoping review: BEME Guide No. 86. Med Teach. :1–17. doi: 10.1080/0142159X.2023.2241621.

Dumić-Čule I, Orešković T, Brkljačić B, Kujundžić Tiljak M, Orešković S. 2020. The importance of introducing artificial intelligence to the medical curriculum–assessing practitioners' perspectives. Croat Med J. 61(5):457–464. doi: 10.3325/cmj.2020.61.457.

Duong MT, Rauschecker AM, Rudie JD, Chen PH, Cook TS, Bryan RN, Mohan S. 2019. Artificial intelligence for precision education in radiology. Br J Radiol. 92(1103):20190389. doi: 10.1259/bjr.20190389.

Ebina K, Abe T, Hotta K, Higuchi M, Furumido J, Iwahara N, Kon M, Miyaji K, Shibuya S, Lingbo Y, et al. 2022. Objective evaluation of laparoscopic surgical skills in wet lab training based on motion analysis and machine learning. Langenbecks Arch Surg. 407(5):2123–2132. doi: 10.1007/s00423-022-02505-9.

Ejaz H, McGrath H, Wong BL, Guise A, Vercauteren T, Shapey J. 2022. Artificial intelligence and medical education: a global mixed-methods study of medical students' perspectives. Digit Health. 8: 20552076221089099. doi: 10.1177/20552076221089099.

El Saadawi GM, Tseytlin E, Legowski E, Jukic D, Castine M, Fine J, Gormley R, Crowley RS. 2008. A natural language intelligent tutoring system for training pathologists: implementation and evaluation. Adv Health Sci Educ Theory Pract. 13(5):709–722. doi: 10.1007/s10459-007-9081-3.

Epstein RH, Dexter F. 2023. Variability in large language models' responses to medical licensing and certification examinations. comment on "how does ChatGPT perform on the United States medical licensing examination? the implications of large language models for medical education and knowledge assessment". JMIR Med Educ. 9:e48305. doi: 10.2196/48305.

Eysenbach G. 2023. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for publications. JMIR Med Educ. 9(1):e46885. doi: 10.2196/46885.

Fan KY, Hu R, Singla R. 2020. Introductory machine learning for medical students: a pilot. Med Educ. 54(11):1042–1043. doi: 10.1111/medu.14318.

Fazlollahi AM, Bakhaidar M, Alsayegh A, Yilmaz R, Winkler-Schwartz A, Mirchi N, Langleben I, Ledwos N, Sabbagh AJ, Bajunaid K, et al. 2022. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. JAMA Netw Open. 5(2):e2149008. doi: 10.1001/jamanetworkopen.2021.49008.

Fekri P, Dargahi J, Zadeh M. 2020. Deep learning-based haptic guidance for surgical skills transfer. Front Robot AI. 7:586707. doi: 10.3389/frobt.2020.586707.

Feng S, Shen Y. 2023. ChatGPT and the future of medical education. Acad Med. 98(8):868. doi: 10.1097/acm.0000000000005242.

Fischetti C, Bhatter P, Frisch E, Sidhu A, Helmy M, Lungren M, Duhaime E. 2022. The evolving importance of artificial intelligence and radiology in medical trainee education. Acad Radiol. 29 Suppl 5:S70–S75. doi: 10.1016/j.acra.2021.03.023.

Friederichs H, Friederichs WJ, März M. 2023. ChatGPT in medical school: how successful is AI in progress testing? Med Educ Online. 28(1):2220920. doi: 10.1080/10872981.2023.2220920.

Frize M, Frasson C. 2000. Decision-support and intelligent tutoring systems in medical education. Clin Investig Med. 23(4):266–269.

Furlan R, Gatti M, Menè R, Shiffer D, Marchiori C, Giaj Levra A, Saturnino V, Brunetta E, Dipaola F. 2021. A natural language processing–based virtual patient simulator and intelligent tutoring system for the clinical diagnostic process: simulator development and case study. JMIR Med Inform. 9(4):e24073. doi: 10.2196/24073.

Furlan R, Gatti M, Mene R, Shiffer D, Marchiori C, Giaj Levra A, Saturnino V, Brunetta E, Dipaola F. 2022. Learning analytics applied to clinical diagnostic reasoning using a natural language processing–based virtual patient simulator: case study. JMIR Med Educ. 8(1):e24372. doi: 10.2196/24372.

Gao Y, Kruger U, Intes X, Schwaitzberg S, De S. 2020. A machine learning approach to predict surgical learning curves. Surgery. 167(2): 321–327. doi: 10.1016/j.surg.2019.10.008.

Gencer A, Aydin S. 2023. Can ChatGPT pass the thoracic surgery exam? Am J Med Sci. 366(4):291–295. doi: 10.1016/j.amjms.2023.08.001.

Gendia A. 2022. Cloud based AI-driven video analytics (CAVs) in laparoscopic surgery: a step closer to a virtual portfolio. Cureus. 14(9): e29087. doi: 10.7759/cureus.29087.

Giannos P. 2023. Evaluating the limits of AI in medical specialisation: chatGPT's performance on the UK neurology specialty certificate examination. BMJ Neurol Open. 5(1):e000451. doi: 10.1136/bmjno-2023-000451.

Gibbons C, Richards S, Valderas JM, Campbell J. 2017. Supervised machine learning algorithms can classify open-text feedback of doctor performance with human-level accuracy. J Med Internet Res. 19(3):e6533. doi: 10.2196/jmir.6533.

Gillissen A, Kochanek T, Zupanic M, Ehlers J. 2022. Medical students' perceptions towards digitization and artificial intelligence: a mixed-methods study. Healthcare. 10(4):723. doi: 10.3390/healthcare10040723.

Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. 2023. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 9(1):e45312. doi: 10.2196/45312.

Gong B, Nugent JP, Guest W, Parker W, Chang PJ, Khosa F, Nicolaou S. 2019. Influence of artificial intelligence on Canadian medical students' preference for radiology specialty: a national survey study. Acad Radiol. 26(4):566–577. doi: 10.1016/j.acra.2018.10.007.

Gopinath RA, Churiwala S. 2019. Machine learning definition and basics. In: An introduction to machine learning. Cham (Switzerland): Springer Nature. .

Gordon M, Gibbs T. 2014. STORIES statement: publication standards for healthcare education evidence synthesis. BMC Med. 12(1):1–9. doi: 10.1186/s12916-014-0143-0.

Gordon M, Patricio M, Horne L, Muston A, Alston S, Pammi M, Thammasitboon S, Park S, Pawlikowska T, Rees E, et al. 2020. Developments in medical education in response to the COVID-19 pandemic: a rapid BEME systematic review: BEME Guide No. 63. Med Teach. 42(11):1202–1215. doi: 10.1080/0142159X.2020.1807484.

Gorospe-Sarasúa L, Muñoz-Olmedo JM, Sendra-Portero F, de Luis-García R. 2022. Challenges of radiology education in the era of artificial intelligence. Radiologia. 64(1):54–59. doi: 10.1016/j.rxeng.2020.10.012.

Gowda V, Jordan SG, Awan OA. 2022. Artificial intelligence in radiology education: a longitudinal approach. Acad Radiol. 29(5):788–790. doi: 10.1016/j.acra.2021.08.022.

Graham A. 2023. ChatGPT and other AI tools put students at risk of plagiarism allegations, MDU warns. BMJ. 381:1133. doi: 10.1136/bmj.p1133.

Grunhut J, Marques O, Wyatt AT. 2022. Needs, challenges, and applications of artificial intelligence in medical education curriculum. JMIR Med Educ. 8(2):e35587. doi: 10.2196/35587.

Grunhut J, Wyatt AT, Marques O. 2021. Educating future physicians in artificial intelligence (AI): an integrative review and proposed changes. J Med Educ Curric Dev. 8:23821205211036836. doi: 10.1177/23821205211036836.

Guerra GA, Hofmann H, Sobhani S, Hofmann G, Gomez D, Soroudi D, Hopkins BS, Dallas J, Pangal DJ, Cheok S, et al. 2023. GPT-4 artificial intelligence model outperforms ChatGPT, medical students, and neurosurgery residents on neurosurgery written board-like questions. World Neurosurg. 179:e160–e165. doi: 10.1016/j.wneu.2023.08.042.

Guigue PA, Meyer R, Thivolle-Lioux G, Brezinov Y, Levin G. 2023. Performance of ChatGPT in French language parcours d'accès spécifique santé test and in OBGYN. Int J Gynaecol Obstet. doi: 10.1002/ijgo.15083.

Guo AA, Li J. 2023. Harnessing the power of ChatGPT in medical education. Med Teach. 45(9):1063. doi: 10.1080/0142159x.2023.2198094.

Hakami KM, Alameer M, Jaawna E, Sudi A, Bahkali B, Mohammed A, Hakami A, Mahfouz MS, Alhazmi AH, Dhayihi TM. 2023. The impact of artificial intelligence on the preference of radiology as a future specialty among medical students at Jazan University, Saudi Arabia:

a cross-sectional study. Cureus. 15(7):e41840. doi: 10.7759/cureus.41840.
Hamdy H, Sreedharan J, Rotgans JI, Zary N, Bahous SA, Venkatramana M, AbdelFattah Elzayat E, Lamba P, Sebastian SK, Momen NK. 2021. Virtual clinical encounter examination (VICEE): a novel approach for assessing medical students' non-psychomotor clinical competency. Med Teach. 43(10):1203–1209. doi: 10.1080/0142159x.2021.1935828.
Hammick M, Dornan T, Steinert Y. 2010. Conducting a best evidence systematic review. Part 1: from idea to data coding. BEME Guide No. 13. Med Teach. 32(1):3–15. doi: 10.3109/01421590903414245.
Hamza D, Hauer K, Oswald A, van Melle E, Ladak Z, Zuna I, Assefa M, Pelletier G, Sebastianski M, Keto-Lambert D, et al. 2023. Making sense of competency-based medical education (CBME) literary conversations: a BEME scoping review: BEME Guide No. 78. Med Teach. 45(8):802–815. doi: 10.1080/0142159X.2023.2168525.
Han R, Yu W, Chen H, Chen Y. 2022. Using artificial intelligence reading label system in diabetic retinopathy grading training of junior ophthalmology residents and medical students. BMC Med Educ. 22(1):258. doi: 10.1186/s12909-022-03272-3.
Hanlon CD, Frosch EM, Shochet RB, Buckingham Shum SJ, Gibson A, Goldberg HR. 2021. Recognizing reflection: computer-assisted analysis of first year medical students' reflective writing. Med Sci Educ. 31(1):109–116. doi: 10.1007/s40670-020-01132-7.
Haruna-Cooper L, Rashid MA. 2023. GPT-4: the future of artificial intelligence in medical school assessments. J R Soc Med. 116(6):218–219. doi: 10.1177/01410768231181251.
Hashimoto DA, Johnson KB. 2023. The use of artificial intelligence tools to prepare medical school applications. Acad Med. 98(9):978–982. doi: 10.1097/acm.0000000000005309.
Hathaway QA, Hogg JP, Lakhani DA. 2022. Need for medical student education in emerging technologies and artificial intelligence: fostering enthusiasm, rather than flight, from specialties most affected by emerging technologies. Acad Radiol. 30(8):1770–1771. doi: 10.1016/j.acra.2022.11.018.
Hayasaka Y, Fujikura T, Kashimura M. 2018. Expectations for the next generation of simulated patients born from thoughtful anticipation of artificial intelligence-equipped robot. J Nippon Med Sch. 85(6):347–349. doi: 10.1272/jnms.jnms.2018_85-57.
Heath JK, Clancy CB, Pluta W, Weissman GE, Anderson U, Kogan JR, Dine CJ, Shea JA. 2023. Natural language processing of learners' evaluations of attendings to identify professionalism lapses. Eval Health Prof. 46(3):225–232. doi: 10.1177/01632787231158128.
Hedderich DM, Keicher M, Wiestler B, Gruber MJ, Burwinkel H, Hinterwimmer F, Czempiel T, Spiro JE, Pinto dos Santos D, Heim D, et al. 2021. AI for doctors—a course to educate medical professionals in artificial intelligence for medical imaging. Healthcare. 9(10):1278. doi: 10.3390/healthcare9101278.
Heng JJ, Teo DB, Tan LF. 2023. The impact of chat generative pre-trained transformer (ChatGPT) on medical education. Postgrad Med J. 99(1176):1125–1127. doi: 10.1093/postmj/qgad058.
Henning PA, Henning J, Glück K. 2021. Artificial intelligence: its future in the health sector and its role for medical education. J Eur CME. 10(1):2014099. doi: 10.1080/21614083.2021.2014099.
Hu R, Fan KY, Pandey P, Hu Z, Yau O, Teng M, Wang P, Li A, Ashraf M, Singla R. 2022. Insights from teaching artificial intelligence to medical students in Canada. Commun Med. 2(1):63. doi: 10.1038/s43856-022-00125-4.
Hu R, Rizwan A, Hu Z, Li T, Chung AD, Kwan BY. 2023. An artificial intelligence training workshop for diagnostic radiology residents. Radiol Artif Intell. 5(2):e220170. doi: 10.1148/ryai.220170.
Huh S. 2023. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. J Educ Eval Health Prof. 20(1):1. doi: 10.3352/jeehp.2023.20.1.
Humar P, Asaad M, Bengur FB, Nguyen V. 2023. ChatGPT is equivalent to first year plastic surgery residents: evaluation of ChatGPT on the plastic surgery in-service exam. Aesthetic Surg J. doi: 10.1093/asj/sjad130.
Huston JC, Kaminski N. 2023. A picture worth a thousand words, created with one sentence: using artificial intelligence–created art to enhance medical education. ATS Sch. doi: 10.34197/ats-scholar.2022-0141PS.
Im C. 2022. Medical student's artificial intelligence education and research experiences. Korean J Med Educ. 34(4):341–344. doi: 10.3946/kjme.2022.243.
Imran N, Jawaid M. 2020. Artificial intelligence in medical education: are we ready for it? Pak J Med Sci. 36(5):857–859. doi: 10.12669/pjms.36.5.3042.
Iroju O, Olaleke J. 2015. A systematic review of natural language processing in healthcare. IJITCS. 7(8):44–50. doi: 10.5815/ijitcs.2015.08.07.
Islam M, Hasan M, Wang X, Germack H, Noor-E-Alam M. 2018. A systematic review on healthcare analytics: application and theoretical perspective of data mining. Healthcare. 6(2):54. doi: 10.3390/healthcare6020054.
James CA, Wheelock KM, Woolliscroft JO. 2021. Machine learning: the next paradigm shift in medical education. Acad Med. 96(7):954–957. doi: 10.1097/acm.0000000000003943.
Jani KH, Jones KA, Jones GW, Amiel J, Barron B, Elhadad N. 2020. Machine learning to extract communication and history-taking skills in OSCE transcripts. Med Educ. 54(12):1159–1170. doi: 10.1111/medu.14347.
Jha N, Shankar PR, Al-Betar MA, Mukhia R, Hada K, Palaian S. 2022. Undergraduate medical students' and interns' knowledge and perception of artificial intelligence in medicine. Adv Med Educ Pract. 13:927–937. doi: 10.2147/amep.s368519.
Jin D, Pan E, Oufattole N, Weng WH, Fang H, Szolovits P. 2021. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. Appl Sci. 11(14):6421. doi: 10.48550/arXiv.2009.13081.
Johnsson V, Søndergaard MB, Kulasegaram K, Sundberg K, Tiblad E, Herling L, Petersen OB, Tolsgaard MG. 2023. Validity evidence supporting clinical skills assessment by artificial intelligence compared with trained clinician raters. Med Educ. 58(1):105–117. doi: 10.1111/medu.15190.
Joly-Chevrier M, Nguyen AX, Lesko-Krleza M, Lefrançois P. 2023. Performance of ChatGPT on a practice dermatology board certification examination. J Cutan Med Surg. 27(4):407–409. doi: 10.1177/12034754231188437.
Karabacak M, Ozkara BB, Margetis K, Wintermark M, Bisdas S. 2023. The advent of generative language models in medical education. JMIR Med Educ. 9:e48163. doi: 10.2196/48163.
Karaca O, Çalışkan SA, Demir K. 2021. Medical artificial intelligence readiness scale for medical students (MAIRS-MS)–development, validity and reliability study. BMC Med Educ. 21(1):112. doi: 10.1186/s12909-021-02546-6.
Katznelson G, Gerke S. 2021. The need for health AI ethics in medical school education. Adv Health Sci Educ Theory Pract. 26(4):1447–1458. doi: 10.1007/s10459-021-10040-3.
Kaur A, Singh S, Chandan JS, Robbins T, Patel V. 2021. Qualitative exploration of digital chatbot use in medical education: a pilot study. Digit Health. 7:20552076211038151. doi: 10.1177/20552076211038151.
Kawka M, Gall TM, Fang C, Liu R, Jiao LR. 2022. Intraoperative video analysis and machine learning models will change the future of surgical training. Intell Surg. 1:13–15. doi: 10.1016/j.isurg.2021.03.001.
Keir G, Hu W, Filippi CG, Ellenbogen L, Woldenberg R. 2023. Using artificial intelligence in medical school admissions screening to decrease inter-and intra-observer variability. JAMIA Open. 6(1):ooad011. doi: 10.1093/jamiaopen/ooad011.
Khan RA, Jawaid M, Khan AR, Sajjad M. 2023. ChatGPT-reshaping medical education and clinical management. Pak J Med Sci. 39(2):605–607. doi: 10.12669/pjms.39.2.7653.
Khumrin P, Ryan A, Judd T, Verspoor K. 2018 Jan 31. Diagnostic machine learning models for acute abdominal pain: towards an e-learning tool for medical students. Stud Health Technol Inform. 245:447–451. doi: 10.3233/978-1-61499-830-3-447.
Kibble J, Plochocki J. 2023. Comparing machine learning models and human raters when ranking medical student performance evaluations. J Grad Med Educ. 15(4):488–493. doi: 10.4300/jgme-d-22-00678.1.
King AJ, Kahn JM, Brant EB, Cooper GF, Mowery DL. 2022. Initial development of an automated platform for assessing trainee performance on case presentations. ATS Sch. 3(4):548–560. doi: 10.34197/ats-scholar.2022-0010oc.
Kintsch W. 2002. The potential of latent semantic analysis for machine grading of clinical case summaries. J Biomed Inform. 35(1):3–7. doi: 10.1016/s1532-0464.(02)00004-7
Kiyasseh D, Laca J, Haque TF, Miles BJ, Wagner C, Donoho DA, Anandkumar A, Hung AJ. 2023. A multi-institutional study using artificial intelligence to provide reliable and fair feedback to surgeons. Commun Med. 3(1):42. doi: 10.1038/s43856-023-00263-3.

Kiyasseh D, Laca J, Haque TF, Otiato M, Miles BJ, Wagner C, Donoho DA, Trinh QD, Anandkumar A, Hung AJ. 2023. Human visual explanations mitigate bias in AI-based assessment of surgeon skills. Npj Digit Med. 6(1):54. doi: 10.1038/s41746-023-00766-2.

Koga S. 2023. The potential of ChatGPT in medical education: focusing on USMLE preparation. Ann Biomed Eng. 51(10):2123–2124. doi: 10.1007/s10439-023-03253-7.

Kolachalama VB, Garg PS. 2018. Machine learning and medical education. NPJ Digit Med. 1(1):54. doi: 10.1038/s41746-018-0061-1.

Koljonen V. 2023. What could we make of AI in plastic surgery education. J Plast Reconstr Aesthet Surg. 81:94–96. doi: 10.1016/j.bjps.2023.04.055.

Krive J, Isola M, Chang L, Patel T, Anderson M, Sreedhar R. 2023. Grounded in reality: artificial intelligence in medical education. JAMIA Open. 6(2):ooad037. doi: 10.1093/jamiaopen/ooad037.

Kumah-Crystal Y, Mankowitz S, Embi P, Lehmann CU. 2023. ChatGPT and the clinical informatics board examination: the end of unproctored maintenance of certification? J Am Med Inform Assoc. 30(9):1558–1560. doi: 10.1093/jamia/ocad104.

Kundu S. 2021. How will artificial intelligence change medical training? Commun Med. 1(1):8. doi: 10.1038/s43856-021-00003-5.

Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, et al. 2023. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2(2):e0000198. doi: 10.1371/journal.pdig.0000198.

Lam A, Lam L, Blacketer C, Parnis R, Franke K, Wagner M, Wang D, Tan Y, Oakden-Rayner L, Gallagher S, et al. 2022. Professionalism and clinical short answer question marking with machine learning. Intern Med J. 52(7):1268–1271. doi: 10.1111/imj.15839.

Lebhar MS, Velazquez A, Goza S, Hoppe IC. 2023. Dr. ChatGPT: utilizing artificial intelligence in surgical education. Cleft Palate Craniofac J. 10556656231193966:10556656231193966. doi: 10.1177/10556656231193966.

Ledwos N, Mirchi N, Yilmaz R, Winkler-Schwartz A, Sawni A, Fazlollahi AM, Bissonnette V, Bajunaid K, Sabbagh AJ, Del Maestro RF. 2022. Assessment of learning curves on a simulated neurosurgical task using metrics selected by artificial intelligence. J Neurosurg. 137(4):1–12. doi: 10.3171/2021.12.jns211563.

Lee D, Yu HW, Kwon H, Kong HJ, Lee KE, Kim HC. 2020. Evaluation of surgical skills during robotic surgery by deep learning-based multiple surgical instrument tracking in training and actual operations. J Clin Med. 9(6):1964. doi: 10.3390/jcm9061964.

Lee H. 2023. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ. doi: 10.1002/ase.2270.

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

Levac D, Colquhoun H, O'Brien KK. 2010. Scoping studies: advancing the methodology. Implement Sci. 5(1):69. doi: 10.1186/1748-5908-5-69.

Li S, Miles K, George R, Ertubey C, Pype P, Liu J. 2023. A critical review of cultural competence frameworks and models in medical and health professional education: a meta-ethnographic synthesis: BEME Guide No. 79. Med Teach. 45(10):1085–1107. doi: 10.1080/0142159X.2023.2174419.

Li X, Jiang MY, Jong MS, Zhang X, Chai CS. 2022. Understanding medical students' perceptions of and behavioral intentions toward learning artificial intelligence: a survey study. Int J Environ Res Public Health. 19(14):8733. doi: 10.3390/ijerph19148733.

Li YS, Lam CS, See C. 2021. Using a machine learning architecture to create an AI-powered chatbot for anatomy education. Med Sci Educ. 31(6):1729–1730. doi: 10.1007/s40670-021-01405-9.

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

Lillehaug SI, Lajoie SP. 1998. AI in medical education—another grand challenge for medical informatics. Artif Intell Med. 12(3):197–225. doi: 10.1016/s0933-3657.(97)00054-7

Lindqwister AL, Hassanpour S, Levy J, Sin JM. 2022. AI-RADS: successes and challenges of a novel artificial intelligence curriculum for radiologists across different delivery formats. Front Med Technol. 4:1007708. doi: 10.3389/fmedt.2022.1007708.

Liu DS, Sawyer J, Luna A, Aoun J, Wang J, Boachie L, Halabi S, Joe B. 2022. Perceptions of US medical students on artificial intelligence in medicine: mixed methods survey study. JMIR Med Educ. 8(4):e38325. doi: 10.2196/38325.

Loh E. 2023. ChatGPT and generative AI chatbots: challenges and opportunities for science, medicine and medical leaders. BMJ Lead. 0:1–4. doi: 10.1136/leader-2023-000797.

Lower K, Seth I, Lim B, Seth N. 2023. ChatGPT-4: transforming medical education and addressing clinical exposure challenges in the post-pandemic era. Indian J Orthop. 57(9):1527–1544. doi: 10.1007/s43465-023-00967-7.

Lu KJ, Meaney C, Guo E, Leung FH. 2023. Evaluating the applicability of existing lexicon-based sentiment analysis techniques on family medicine resident feedback field notes: retrospective cohort study. JMIR Med Educ. 9(1):e41953. doi: 10.2196/41953.

Maicher KR, Zimmerman L, Wilcox B, Liston B, Cronau H, Macerollo A, Jin L, Jaffe E, White M, Fosler-Lussier E, et al. 2019. Using virtual standardized patients to accurately assess information gathering skills in medical students. Med Teach. 41(9):1053–1059. doi: 10.1080/0142159x.2019.1616683.

Maini B, Maini E. 2021. Artificial intelligence in medical education. Indian Pediatr. 58(5):496–497. [No DOI] doi: 10.1007/s13312-021-2228-0.

Masters K. 2019. Artificial intelligence in medical education. Med Teach. 41(9):976–980. doi: 10.1080/0142159X.2019.1595557.

Masters K. 2023a. Ethical use of artificial intelligence in health professions education: AMEE Guide No.158. Med Teach. 45(6):574–584. doi: 10.1080/0142159X.2023.2186203.

Masters K. 2023b. Medical Teacher's first ChatGPT's referencing hallucinations: lessons for editors, reviewers, and teachers. Med Teach. 45(7):673–675. doi: 10.1080/0142159X.2023.2208731.

Mastour H, Dehghani T, Moradi E, Eslami S. 2023. Early prediction of medical students' performance in high-stakes examinations using machine learning approaches. Heliyon. 9(7):e18248. doi: 10.1016/j.heliyon.2023.e18248.

Mbakwe AB, Lourentzou I, Celi LA, Mechanic OJ, Dagan A. 2023. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. PLOS Digit Health. 2(2):e0000205. doi: 10.1371/journal.pdig.0000205.

McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. 2020. What do medical students actually need to know about artificial intelligence? NPJ Digit Med. 3(1):86. doi: 10.1038/s41746-020-0294-7.

Mehta N, Harish V, Bilimoria K, Morgado F, Ginsburg S, Law M, Das S. 2021. Knowledge of and attitudes on artificial intelligence in healthcare: a provincial survey study of medical students [version 1]. MedEdPublish. 10(1). doi: 10.15694/mep.2021.000075.1.

Menon A, Gaglani S, Haynes MR, Tackett S. 2017. Using "big data" to guide implementation of a web and mobile adaptive learning platform for medical students. Med Teach. 39(9):975–980. doi: 10.1080/0142159x.2017.1324949.

Meo SA, Al-Masri AA, Alotaibi M, Meo MZ, Meo MO. 2023. ChatGPT knowledge evaluation in basic and clinical medical sciences: multiple choice question examination-based performance. Healthcare. 11(14):2046. doi: 10.3390/healthcare11142046.

Merritt C, Glisson M, Dewan M, Klein M, Zackoff M. 2022. Implementation and evaluation of an artificial intelligence driven simulation to improve resident communication with primary care providers. Acad Pediatr. 22(3):503–505. doi: 10.1016/j.acap.2021.12.013.

Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. 2020. The virtual operative assistant: an explainable artificial intelligence tool for simulation-based training in surgery and medicine. PLoS One. 15(2):e0229596. doi: 10.1371/journal.pone.0229596.

Mogali SR. 2023. Initial impressions of ChatGPT for anatomy education. Anat Sci Educ. doi: 10.1002/ase.2261.

Moglia A, Morelli L, D'Ischia R, Fatucchi LM, Pucci V, Berchiolli R, Ferrari M, Cuschieri A. 2022. Ensemble deep learning for the prediction of proficiency at a virtual simulator for robot-assisted surgery. Surg Endosc. 36(9):6473–6479. doi: 10.1007/s00464-021-08999-6.

Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A. 2023. Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. Med Educ Online. 28(1):2182659. doi: 10.1080/10872981.2023.2182659.

Moritz S, Romeike B, Stosch C, Tolks D. 2023. Generative AI (gAI) in medical education: chat-GPT and co. GMS J Med Educ. 40(4):Doc54. doi: 10.3205/zma001636.

Morreel S, Mathysen D, Verhoeven V. 2023. Aye, AI! ChatGPT passes multiple-choice family medicine exam. Med Teach. 45(6):665–666. doi: 10.1080/0142159x.2023.2187684.

Mosch L, Agha-Mir-Salim L, Sarica MM, Balzer F, Poncette AS. 2022. Artificial intelligence in undergraduate medical education. Challenges of trustable AI and added-value on health [e-book]. Amsterdam: IOS Press; p. 821–822. doi: 10.3233/shti220597.

Muntean GA, Groza A, Marginean A, Slavescu RR, Steiu MG, Muntean V, Nicoara SD. 2023. Artificial intelligence for personalised ophthalmology residency training. J Clin Med. 12(5):1825. doi: 10.3390/jcm12051825.

Nagaraj MB, Namazi B, Sankaranarayanan G, Scott DJ. 2023. Developing artificial intelligence models for medical student suturing and knot-tying video-based assessment and coaching. Surg Endosc. 37(1):402–411. doi: 10.1007/s00464-022-09509-y.

Nagy M, Radakovich N, Nazha A. 2022. Why machine learning should be taught in medical schools. Med Sci Educ. 32(2):529–532. doi: 10.1007/s40670-022-01502-3.

Natheir S, Christie S, Yilmaz R, Winkler-Schwartz A, Bajunaid K, Sabbagh AJ, Werthner P, Fares J, Azarnoush H, Del Maestro R. 2023. Utilizing artificial intelligence and electroencephalography to assess expertise on a simulated neurosurgical task. Comput Biol Med. 152: 106286. doi: 10.1016/j.compbiomed.2022.106286.

Naylor C. 2018. On the prospects for a (deep) learning health care system. JAMA. 320(11):1099–1100. doi: 10.1001/jama.2018.11103.

Neves SE, Chen MJ, Ku CM, Karan S, DiLorenzo AN, Schell RM, Lee DE, Diachun CA, Jones SB, Mitchell JD. 2021. Using machine learning to evaluate attending feedback on resident performance. Anesth Analg. 132(2):545–555. doi: 10.1213/ane.0000000000005265.

Ngo B, Nguyen D, vanSonnenberg E. 2022. The cases for and against artificial intelligence in the medical school curriculum. Radiol Artif Intell. 4(5):e220074. doi: 10.1148/ryai.220074.

Nguyen GK, Shetty AS. 2018. Artificial intelligence and machine learning: opportunities for radiologists in training. J Am Coll Radiol. 15(9):1320–1321. doi: 10.1016/j.jacr.2018.05.024.

Oberg K. 1960. Cultural shock: adjustment to new cultural environments. Pract Anthropol. os-7(4):177–182. doi: 10.1177/009182966000700405.

Oh N, Choi GS, Lee WY. 2023. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. Ann Surg Treat Res. 104(5):269–273. doi: 10.4174/astr.2023.104.5.269.

Oliven A, Nave R, Gilad D, Barch A. 2011. Implementation of a web-based interactive virtual patient case simulation as a training and assessment tool for medical students. Amsterdam: IOS Press; p. 233–237. doi: 10.3233/978-1-60750-806-9-233.

Ooi SK, Makmur A, Soon AY, Fook-Chong S, Liew C, Sia SY, Ting YH, Lim CY. 2021. Attitudes toward artificial intelligence in radiology with learner needs assessment within radiology residency programmes: a national multi-programme survey. Singapore Med J. 62(3):126–134. doi: 10.11622/smedj.2019141.

Oquendo YA, Riddle EW, Hiller D, Blinman TA, Kuchenbecker KJ. 2018. Automatically rating trainee skill at a pediatric laparoscopic suturing task. Surg Endosc. 32(4):1840–1857. doi: 10.1007/s00464-017-5873-6.

Ossa LA, Rost M, Lorenzini G, Shaw DM, Elger BS. 2023. A smarter perspective: learning with and from AI-cases. Artif Intell Med. 135: 102458. doi: 10.1016/j.artmed.2022.102458.

Ötleş E, James CA, Lomis KD, Woolliscroft JO. 2022. Teaching artificial intelligence as a fundamental toolset of medicine. Cell Rep Med. 3(12):100824. doi: 10.1016/j.xcrm.2022.100824.

Ötleş E, Kendrick DE, Solano QP, Schuller M, Ahle SL, Eskender MH, Carnes E, George BC. 2021. Using natural language processing to automatically assess feedback quality: findings from 3 surgical residencies. Acad Med. 96(10):1457–1460. doi: 10.1097/acm.0000000000004153.

Panthier C, Gatinel D. 2023. Success of ChatGPT, an AI language model, in taking the French language version of the European board of ophthalmology examination: a novel approach to medical knowledge assessment. J Fr Ophtalmol. 46(7):706–711. doi: 10.1016/j.jfo.2023.05.006.

Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. 2019. Introducing artificial intelligence training in medical education. JMIR Med Educ. 5(2):e16048. doi: 10.2196/16048.

Park CJ, Paul HY, Siegel EL. 2021. Medical student perspectives on the impact of artificial intelligence on the practice of medicine. Curr Probl Diagn Radiol. 50(5):614–619. doi: 10.1067/j.cpradiol.2020.06.011.

Park SH, Do KH, Kim S, Park JH, Lim YS. 2019. What should medical students know about artificial intelligence in medicine? J Educ Eval Health Prof. 16:18. doi: 10.3352/jeehp.2019.16.18.

Pereira DS, Falcão F, Nunes A, Santos N, Costa P, Pêgo JM. 2023. Designing and building OSCEBot® for virtual OSCE–performance evaluation. Med Educ Online. 28(1):2228550. doi: 10.1080/10872981.2023.2228550.

Persad A, Stroulia E, Forgie S. 2016. A novel approach to virtual patient simulation using natural language processing. Med Educ. 50(11): 1162–1163. doi: 10.1111/medu.13197.

Peters M, Godfrey C, McInerney P, Khalil H, Larsen P, Marnie C, Pollock D, Tricco AC, Munn Z. 2022. Best practice guidance and reporting items for the development of scoping review protocols. JBI Evid Synth. 20(4):953–968. doi: 10.11124/JBIES-21-00242.

Peters M, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, McInerney P, Godfrey C, Hanan K. 2020. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth. 18(10): 2119–2126. doi: 10.11124/JBIES-20-00167.

Pinto dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, Maintz D, Baeßler B. 2019. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol. 29(4):1640–1646. doi: 10.1007/s00330-018-5601-1.

Prasher J. 2021. Artificial intelligence in medical education. Acad Med. 96(9):1229. doi: 10.1097/acm.0000000000004182.

Pucchio A, Eisenhauer EA, Moraes FY. 2021. Medical students need artificial intelligence and machine learning training. Nat Biotechnol. 39(3):388–389. doi: 10.1038/s41587-021-00846-2.

Pucchio A, Rathagirishnan R, Caton N, Gariscsak PJ, Del Papa J, Nabhen JJ, Vo V, Lee W, Moraes FY. 2022. Exploration of exposure to artificial intelligence in undergraduate medical education: a Canadian cross-sectional mixed-methods study. BMC Med Educ. 22(1):815. doi: 10.1186/s12909-022-03896-5.

Puri P, Jha S. 2021. Artificial intelligence, automation, and medical education: lessons from economic history. J Am Coll Radiol. 18(9):1345–1347. doi: 10.1016/j.jacr.2021.05.002.

Qian X, Jingying H, Xian S, Yuqing Z, Lili W, Baorui C, Wei G, Yefeng Z, Qiang Z, Chunyan C, et al. 2022. The effectiveness of artificial intelligence-based automated grading and training system in education of manual detection of diabetic retinopathy. Front Public Health. 10:1025271. doi: 10.3389/fpubh.2022.1025271.

Rahimzadeh V, Kostick-Quenet K, Blumenthal Barby J, McGuire AL. 2023. Ethics education for healthcare professionals in the era of chatGPT and other large language models: do we still need it? Am J Bioeth. 23(10):17–27. doi: 10.1080/15265161.2023.2233358.

Rampton V, Mittelman M, Goldhahn J. 2020. Implications of artificial intelligence for medical education. Lancet Digit Health. 2(3):e111-2–e112. doi: 10.1016/s2589-7500.(20)30023-6

Reeder K, Lee H. 2022. Impact of artificial intelligence on US medical students' choice of radiology. Clin Imaging. 81:67–71. doi: 10.1016/j.clinimag.2021.09.018.

Rees CA, Ryder HF. 2023. Machine learning for the prediction of ranked applicants and matriculants to an internal medicine residency program. Teach Learn Med. 35(3):277–286. doi: 10.1080/10401334.2022.2059664.

Rezazadeh H, Ahmadipour H, Salajegheh M. 2023. Psychometric evaluation of Persian version of medical artificial intelligence readiness scale for medical students. BMC Med Educ. 23(1):527. doi: 10.1186/s12909-023-04516-6.

Rice DR, Rice GN, Baugh C, Cloutier RL. 2022. Correlation of narrative evaluations to clerkship grades using statistical sentiment analysis. Med Sci Educ. 32(6):1397–1403. doi: 10.1007/s40670-022-01654-2.

Richardson ML, Ojeda PI. 2022. A "bumper-car" curriculum for teaching deep learning to radiology residents. Acad Radiol. 29(5):763–770. doi: 10.1016/j.acra.2021.11.016.

Rogers MP, DeSantis AJ, Janjua H, Barry TM, Kuo PC. 2021. The future surgical training paradigm: virtual reality and machine learning in surgical education. Surgery. 169(5):1250–1252. doi: 10.1016/j.surg.2020.09.040.

Rojek AE, Khanna R, Yim JW, Gardner R, Lisker S, Hauer KE, Lucey C, Sarkar U. 2019. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. J Gen Intern Med. 34(5):684–691. doi: 10.1007/s11606-019-04889-9.

Roos J, Kasapovic A, Jansen T, Kaczmarczyk R. 2023. Artificial intelligence in medical education: comparative analysis of ChatGPT, Bing, and medical students in Germany. JMIR Med Educ. 9(1):e46482. doi: 10.2196/46482.

Ruberto AJ, Rodenburg D, Ross K, Sarkar P, Hungler PC, Etemad A, Howes D, Clarke D, McLellan J, Wilson D. 2021. The future of simulation-based medical education: adaptive simulation utilizing a deep multitask neural network. AEM Educ Train. 5(3):e10605. doi: 10.1002/aet2.10605.

Rujano-Balza MA. 2021. Histology classifier app: remote laboratory sessions using artificial neural networks. Med Sci Educ. 31(2):305–307. doi: 10.1007/s40670-021-01206-0.

Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. 2023. The role of large language models in medical education: applications and implications. JMIR Med Educ. 9:e50945. doi: 10.2196/50945.

Salt J, Harik P, Barone MA. 2019. Leveraging natural language processing: toward computer-assisted scoring of patient notes in the USMLE step 2 clinical skills exam. Acad Med. 94(3):314–316. doi: 10.1097/acm.0000000000002558.

Sarraf D, Vasiliu V, Imberman B, Lindeman B. 2021. Use of artificial intelligence for gender bias analysis in letters of recommendation for general surgery residency candidates. Am J Surg. 222(6):1051–1059. doi: 10.1016/j.amjsurg.2021.09.034.

Sassis L, Kefala-Karli P, Sassi M, Zervides C. 2021. Exploring medical students' and faculty's perception on artificial intelligence and robotics. a questionnaire survey. JoAIMS. 2(1-2):76–84. doi: 10.2991/jaims.d.210617.002.

Savage TR. 2021. Artificial intelligence in medical education. Acad Med. 96(9):1230. doi: 10.1097/acm.0000000000004183.

Schaye V, Guzman B, Burk-Rafel J, Marin M, Reinstein I, Kudlowitz D, Miller L, Chun J, Aphinyanaphongs Y. 2022. Development and validation of a machine learning model for automated assessment of resident clinical reasoning documentation. J Gen Intern Med. 37(9):2230–2238. doi: 10.1007/s11606-022-07526-0.

Schneider M, Bartko P, Geller W, Dannenberg V, König A, Binder C, Goliasch G, Hengstenberg C, Binder T. 2021. A machine learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. Int J Cardiovasc Imaging. 37(2):577–586. doi: 10.1007/s10554-020-02046-6.

Seetharaman R. 2023. Revolutionizing medical education: can ChatGPT boost subjective learning and expression? J Med Syst. 47(1):61. doi: 10.1007/s10916-023-01957-w.

Shoja MM, Van de Ridder JM, Rajput V. 2023. The emerging role of generative artificial intelligence in medical education, research, and practice. Cureus. 15(6):e40883. doi: 10.7759/cureus.40883.

Siddiqui A, Zhao Z, Pan C, Rudzicz F, Everett T. 2023. Deep learning model for automated trainee assessment during high-fidelity simulation. Acad Med. 98(11):1274–1277. doi: 10.1097/acm.0000000000005290.

Sinha RK, Roy AD, Kumar N, Mondal H, Sinha R. 2023. Applicability of ChatGPT in assisting to solve higher order problems in pathology. Cureus. 15(2):e35237. doi: 10.7759/cureus.35237.

Sit C, Srinivasan R, Amlani A, Muthuswamy K, Azam A, Monzon L, Poon DS. 2020. Attitudes and perceptions of UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights Imaging. 11(1):14. doi: 10.1186/s13244-019-0830-7.

Sivasubramanian J, Hussain SM, Muthuprakash SV, Periadurai ND, Mohanram K, Surapaneni KM. 2023. Analysing the clinical knowledge of ChatGPT in medical microbiology in the undergraduate medical examination. Indian J Med Microbiol. 45:100380. doi: 10.1016/j.ijmmb.2023.100380.

Skalidis I, Cagnina A, Luangphiphat W, Mahendiran T, Muller O, Abbe E, Fournier S. 2023. ChatGPT takes on the European exam in core cardiology: an artificial intelligence success story? Eur Heart J Digit Health. 4(3):279–281. doi: 10.1093/ehjdh/ztad029.

Solano QP, Hayward L, Chopra Z, Quanstrom K, Kendrick D, Abbott KL, Kunzmann M, Ahle S, Schuller M, Ötleş E, et al. 2021. Natural language processing and assessment of resident feedback quality. J Surg Educ. 78(6):e72–e77. doi: 10.1016/j.jsurg.2021.05.012.

Soong TK, Ho CM. 2021. Artificial intelligence in medical OSCEs: reflections and future developments. Adv Med Educ Pract. 12:167–173. doi: 10.2147/amep.s287926.

Stevens RH, Najafi K. 1992. Can artificial neural networks provide an "expert's" view of medical students performances on computer based simulations? Proc Annu Symp Comput Appl Med Care. 1992:179–183.

Stewart J, Lu J, Gahungu N, Goudie A, Fegan PG, Bennamoun M, Sprivulis P, Dwivedi G. 2023. Western Australian medical students' attitudes towards artificial intelligence in healthcare. PLoS One. 18(8):e0290642. doi: 10.1371/journal.pone.0290642.

Strong E, DiGiammarino A, Weng Y, Basaviah P, Hosamani P, Kumar A, Nevins A, Kugler J, Hom J, Chen J. 2023. Performance of ChatGPT on free-response, clinical reasoning exams. medRxiv. 2023:2023–2003. [This article is a preprint and has not been peer-reviewed.] doi: 10.1101/2023.03.24.23287731.

Subramani M, Jaleel I, Krishna Mohan S. 2023. Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. Adv Physiol Educ. 47(2):270–271. doi: 10.1152/advan.00036.2023.

Suchman K, Garg S, Trindade AJ. 2023. Chat generative pretrained transformer fails the multiple-choice American College of Gastroenterology self-assessment test. Am J Gastroenterol. 118(12):2280–2282. doi: 10.14309/ajg.0000000000002320.

Sugamiya Y, Otani T, Nakadate R, Takanishi A. 2019. Construction of automatic scoring system to support objective evaluation of clinical skills in medical education. Annu Int Conf IEEE Eng Med Biol Soc. 2019:4177–4181. doi: 10.1109/EMBC.2019.8856335.

Swed S, Alibrahim H, Elkalagi NK, Nasif MN, Rais MA, Nashwan AJ, Aljabali A, Elsayed M, Sawaf B, Albuni MK, et al. 2022. Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in Syria: a cross-sectional online survey. Front Artif Intell. 5:1011524. doi: 10.3389/frai.2022.1011524.

Tajmir SH, Alkasab TK. 2018. Toward augmented radiologists: changes in radiology education in the era of machine learning and artificial intelligence. Acad Radiol. 25(6):747–750. doi: 10.1016/j.acra.2018.03.007.

Takagi S, Watari T, Erabi A, Sakaguchi K. 2023. Performance of GPT-3.5 and GPT-4 on the Japanese medical licensing examination: comparison study. JMIR Med Educ. 9(1):e48002. doi: 10.2196/48002.

Tejani AS, Fielding JR, Peshock RM. 2021. What should radiology residency and fellowship training in artificial intelligence include? A trainee's perspective—radiology in training. Radiology. 299(2):E243–E245. doi: 10.1148/radiol.2021204406.

Tejani AS, Elhalawani H, Moy L, Kohli M, Kahn CE. Jr 2023. Artificial Intelligence and radiology education. Radiol Artif Intell. 5(1):e220084. doi: 10.1148/ryai.220084.

Thanawala R, Jesneck J, Shelton J, Rhee R, Seymour NE. 2022. Overcoming systems factors in case logging with artificial intelligence tools. J Surg Educ. 79(4):1024–1030. doi: 10.1016/j.jsurg.2022.01.013.

Tran AQ, Nguyen LH, Nguyen HS, Nguyen CT, Vu LG, Zhang M, Vu TM, Nguyen SH, Tran BX, Latkin CA, et al. 2021. Determinants of intention to use artificial intelligence-based diagnosis support system among prospective physicians. Front Public Health. 9:755644. doi: 10.3389/fpubh.2021.755644.

Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MD, Horsley T, Weeks L, et al. 2018. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 169(7):467–473. https://. doi: 10.7326/M18-0850.

Triola MM, Reinstein I, Marin M, Gillespie C, Abramson S, Grossman RI, Rivera R. 2023. Mar 6 Artificial intelligence screening of medical school applications: development and validation of a machine-learning algorithm. J Assoc Am Med Coll. 98(9):1036–1043. doi: 10.1097/ACM.0000000000005202.

Tsang R. 2023. Practical applications of ChatGPT in undergraduate medical education. J Med Educ Curric Dev. 10:23821205231178449. doi: 10.1177/23821205231178449.

Tschirhart J, Woolsey A, Skinner J, Ahmed K, Fleming C, Kim J, Dave C, Arntfield R. 2023. Introducing medical students to deep learning through image labelling: a new approach to meet calls for greater artificial intelligence fluency among medical trainees. Can Med Educ J. 14(3):113–115. doi: 10.36834/cmej.75074.

Tsopra R, Peiffer-Smadja N, Charlier C, Campeotto F, Lemogne C, Ruszniewski P, Vivien B, Burgun A. 2023. Putting undergraduate medical students in AI-CDSS designers' shoes: an innovative teaching method to develop digital health critical thinking. Int J Med Inform. 171:104980. doi: 10.1016/j.ijmedinf.2022.104980.

Valikodath NG, Al-Khaled T, Cole E, Ting DS, Tu EY, Campbell JP, Chiang MF, Hallak JA, Chan RP. 2021. Evaluation of pediatric

ophthalmologists' perspectives of artificial intelligence in ophthalmology. J Aapos. 25(3):164-e1–164.e5. doi: 10.1016/j.jaapos.2021.01.011.

Valikodath NG, Cole E, Ting DS, Campbell JP, Pasquale LR, Chiang MF, Chan RP. 2021. Impact of artificial intelligence on medical education in ophthalmology. Transl Vis Sci Technol. 10(7):14. doi: 10.1167/tvst.10.7.14.

van de Ridder JM, Shoja MM, Rajput V. 2023. Finding the place of ChatGPT in medical education. Acad Med. 98(8):867. doi: 10.1097/acm.0000000000005254.

van Hoek J, Huber A, Leichtle A, Härmä K, Hilt D, von Tengg-Kobligk H, Heverhagen J, Poellinger A. 2019. A survey on the future of radiology among radiologists, medical students and surgeons: students and surgeons tend to be more skeptical about artificial intelligence and radiologists may fear that other disciplines take over. Eur J Radiol. 121:108742. doi: 10.1016/j.ejrad.2019.108742.

Velez-Florez MC, Ghosh A, Patton D, Sze R, Reid JR, Sotardi S. 2023. Artificial intelligence curriculum needs assessment for a pediatric radiology fellowship program: what, how, and why? Acad Radiol. 30(2):349–358. doi: 10.1016/j.acra.2022.04.026.

Vignesh R, Pradeep P, Balakrishnan P. 2023. A tête-à-tête with ChatGPT on the impact of artificial intelligence in medical education. Med J Malaysia. 78(4):547–549. NO DOI

Wagner G, Raymond L, Paré G. 2023. Understanding prospective physicians' intention to use artificial intelligence in their future medical practice: configurational analysis. JMIR Med Educ. 9:e45631. doi: 10.2196/45631.

Waldman CE, Hermel M, Hermel JA, Allinson F, Pintea MN, Bransky N, Udoh E, Nicholson L, Robinson A, Gonzalez J, et al. 2022. Artificial intelligence in healthcare: a primer for medical education in radiomics. Per Med. 19(5):445–456.,. doi: 10.2217/pme-2022-0014.

Wang M, Sun Z, Jia M, Wang Y, Wang H, Zhu X, Chen L, Ji H. 2022. Intelligent virtual case learning system based on real medical records and natural language processing. BMC Med Inform Decis Mak. 22(1):60. doi: 10.1186/s12911-022-01797-7.

Wang LK, Paidisetty PS, Cano AM. 2023. The next paradigm shift? ChatGPT, artificial intelligence, and medical education. Med Teach. 45(8):925. doi: 10.1080/0142159x.2023.2198663.

Wang X, Gong Z, Wang G, Jia J, Xu Y, Zhao J, Fan Q, Wu S, Hu W, Li X. 2023. Chatgpt performs on the Chinese national medical licensing examination. J Med Syst. 47(1):86. doi: 10.1007/s10916-023-01961-0.

Wang Z, Majewicz Fey A. 2018. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. Int J Comput Assist Radiol Surg. 13(12):1959–1970. doi: 10.1007/s11548-018-1860-1.

Ward TM, Mascagni P, Madani A, Padoy N, Perretta S, Hashimoto DA. 2021. Surgical data science and artificial intelligence for surgical education. J Surg Oncol. 124(2):221–230. doi: 10.1002/jso.26496.

Wartman SA, Combs CD. 2018. Medical education must move from the information age to the age of artificial intelligence. Acad Med. 93(8):1107–1109. doi: 10.1097/acm.0000000000002044.

Wartman SA, Combs CD. 2019. Reimagining medical education in the age of AI. AMA J Ethics. 21(2):146–152. doi: 10.1001/amajethics.2019.146.

Webb JJ. 2023. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. Cureus. 15(5):e38755. doi: 10.7759/cureus.38755.

Weidener L, Fischer M. 2023. Artificial intelligence teaching as part of medical education: qualitative analysis of expert interviews. JMIR Med Educ. 9(1):e46428. doi: 10.2196/46428.

Wiggins WF, Caton MT, Magudia K, Glomski SH, George E, Rosenthal MH, Gaviola GC, Andriole KP. 2020. Preparing radiologists to lead in the era of artificial intelligence: designing and implementing a focused data science pathway for senior radiology residents. Radiol Artif Intell. 2(6):e200057. doi: 10.1148/ryai.2020200057.

Witten AJ, Patel N, Cohen-Gadol A. 2022. Image segmentation of operative neuroanatomy into tissue categories using a machine learning construct and its role in neurosurgical training. Oper Neurosurg. 23(4):279–286. doi: 10.1227/ons.0000000000000322.

Woo CW, Evens MW, Freedman R, Glass M, Shim LS, Zhang Y, Zhou Y, Michael J. 2006. An intelligent tutoring system that generates a natural language dialogue using dynamic multi-level planning. Artif Intell Med. 38(1):25–46. doi: 10.1016/j.artmed.2005.10.004.

Wood EA, Ange BL, Miller DD. 2021. Are we ready to integrate artificial intelligence literacy into medical school curriculum: students and faculty survey. J Med Educ Curric Dev. 8:23821205211024078. doi: 10.1177/23821205211024078.

Xie Y, Seth I, Hunter-Smith DJ, Rozen WM, Seifman MA. 2023. Investigating the impact of innovative AI chatbot on post-pandemic medical education and clinical assistance: a comprehensive analysis. ANZ J Surg. doi: 10.1111/ans.18666.

Yang L, Zheng S, Xu X, Sun Y, Wang X, Li J. 2021. Medical data mining course development in postgraduate medical education: web-based survey and case study. JMIR Med Educ. 7(4):e24027. doi: 10.2196/24027.

Yang W, Hebert D, Kim S, Kang B. 2019. MCRDR knowledge-based 3D dialogue simulation in clinical training and assessment. J Med Syst. 43(7):200. doi: 10.1007/s10916-019-1262-0.

Yi PK, Ray ND, Segall N. 2023. A novel use of an artificially intelligent chatbot and a live, synchronous virtual question-and answer session for fellowship recruitment. BMC Med Educ. 23(1):152. doi: 10.1186/s12909-022-03872-z.

Yilmaz Y, Jurado Nunez A, Ariaeinejad A, Lee M, Sherbino J, Chan TM. 2022. Harnessing natural language processing to support decisions around workplace-based assessment: machine learning study of competency-based medical education. JMIR Med Educ. 8(2):e30537. doi: 10.2196/30537.

Zack T, Dhaliwal G, Geha R, Margaretten M, Murray S, Hong JC. 2023. A clinical reasoning-encoded case library developed through natural language processing. J Gen Intern Med. 38(1):5–11. doi: 10.1007/s11606-022-07758-0.

Zhang R, Pakhomov S, Gladding S, Aylward M, Borman-Shoap E, Melton GB. 2012. Automated assessment of medical training evaluation text. AMIA Annu Symp Proc. 2012:1459–1468.

Zhao S, Xiao X, Zhang X, Meng WL, Soghier L, Hahn JK. 2020. Automated assessment system for neonatal endotracheal intubation using dilated convolutional neural network. Annu Int Conf IEEE Eng Med Biol Soc. 2020:5455–5458. doi: 10.1109/embc44109.2020.9176329.

Zhao S, Zhang X, Jin F, Hahn J. 2021. An auxiliary tasks based framework for automated medical skill assessment with limited data. Conf IEEE Eng Med Biol Soc; p. 1613–1617. doi: 10.1109/embc46164.2021.9630498.

Zhou Y, Jindal-Snape D, Topping K, Todman J. 2008. Theoretical models of culture shock and adaptation in international students in higher education. Stud High Educ. 33(1):63–75. doi: 10.1080/03075070701794833.

Zhong JY, Fischer NL. 2023. Commentary: the desire of medical students to integrate artificial intelligence into medical education: an opinion article. Front Digit Health. 5:1151390. doi: 10.3389/fdgth.2023.1151390.

Zumsteg JM, Junn C. 2023. Will ChatGPT match to your program? Am J Phys Med Rehabil. 102(6):545–547. doi: 10.1097/phm.0000000000002238.