

RESEARCH

Open Access



A systematic review of the impact of artificial intelligence on educational outcomes in health professions education

Eva Feigerlova^{1,2,3,5*} , Hind Hani^{1,2} and Ellie Hothersall-Davies⁴

Abstract

Background Artificial intelligence (AI) has a variety of potential applications in health professions education and assessment; however, measurable educational impacts of AI-based educational strategies on learning outcomes have not been systematically evaluated.

Methods A systematic literature search was conducted using electronic databases (CINAHL Plus, EMBASE, Proquest, Pubmed, Cochrane Library, and Web of Science) to identify studies published until October 1st 2024, analyzing the impact of AI-based tools/interventions in health profession assessment and/or training on educational outcomes. The present analysis follows the PRISMA 2020 statement for systematic reviews and the structured approach to reporting in health care education for evidence synthesis.

Results The final analysis included twelve studies. All were single centers with sample sizes ranging from 4 to 180 participants. Three studies were randomized controlled trials, and seven had a quasi-experimental design. Two studies were observational. The studies had a heterogenous design. Confounding variables were not controlled. None of the studies provided learning objectives or descriptions of the competencies to be achieved. Three studies applied learning theories in the development of AI-powered educational strategies. One study reported the analysis of the authenticity of the learning environment. No study provided information on the impact of feedback activities on learning outcomes. All studies corresponded to Kirkpatrick's second level evaluating technical skills or quantifiable knowledge. No study evaluated more complex tasks, such as the behavior of learners in the workplace. There was insufficient information on training datasets and copyright issues.

Conclusions The results of the analysis show that the current evidence regarding measurable educational outcomes of AI-powered interventions in health professions education is poor. Further studies with a rigorous methodological approach are needed. The present work also highlights that there is no straightforward guide for evaluating the quality of research in AI-based education and suggests a series of criteria that should be considered.

Trial registration Methods and inclusion criteria were defined in advance, specified in a protocol and registered in the OSF registries (<https://osf.io/v5cgp/>). Clinical Trial number: not applicable.

Keywords Artificial intelligence, Health professions, AI-based training and assessment, Learning theories and principles, Educational outcomes

*Correspondence:

Eva Feigerlova

eva.feigerlova@fulbrightmail.org; eva.feigerlova@univ-lorraine.fr

Full list of author information is available at the end of the article



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

Introduction

Artificial intelligence (AI) has a variety of potential applications in health professions education, such as personalized learning or performance assessment [1–3]. A large amount of information can be used to evaluate the performance and progress of students using AI-based learning analytics tools [4–6]. Although AI has produced promising results, it also raises important doubts and concerns [1, 7]. Variable accuracies of AI-based educational models have been reported [4, 6]. For example, large language models (LLMs) such as the generative pretrained transformer can yield variable responses irrespective of the prompts used [8, 9]. Educators must be aware of the potential inaccuracy of generated data when deciding what teaching strategy to adopt. AI can also lead to challenges in terms of the standardization of teaching and create disparities in access to AI-based technologies [10]. Many AI-based tools used in clinical settings are designed to provide assistance to clinicians rather than increase their knowledge [11, 12]. Literature data also indicate some situations, which may negatively impact the development of critical thinking and/or clinical reasoning of learners, such as an overreliance on AI-tool generated responses, a lack of original thoughts of learners, or a lack of contextual understanding of the AI-generated responses [13]. In fact, medical expertise involves not only knowledge of pathophysiology but also cognitive reasoning strategies in specific clinical situations [14].

When developing AI-based pedagogical strategies, one must keep in mind how to measure the attainment of learning outcomes. Student learning can be measured on different levels, such as a model proposed by Kirkpatrick [15], which enables the evaluation of training effectiveness from the reactions of participants to their performance in a workplace-based environment. Nevertheless, to evaluate AI-powered educational interventions in health professions education, several other aspects must be considered, such as the availability of users' data, the datasets used to train machine learning (ML) models or the pertinence of a given model to the educational needs of a specific institution [16]. Another important element that must be taken into account is the risk related to model bias, such as inaccuracies in datasets or anomalous conclusions of algorithms [1, 17, 18]. Generative AI can be a source of academic misconduct and misuse of AI-generated data [1, 19]. This raises legal and ethical questions such as the protection of personal data [7].

The literature available in this research field [2, 7, 13, 20–23] has reported a broad potential utility of AI for training purposes in health professions education but there has been no deep analysis of learning impact. In the scoping review, Gordon et al. [24] synthesized 278 publications showing a large spectrum of stages,

specialties and uses of AI in medical education. Most studies reflected early adaptation phases. Only a few papers reported the application of AI for longitudinal educational changes. Lee et al. [25] reviewed 22 publications regarding utilization of AI in the undergraduate medical education. There was an important heterogeneity of studies and a weak concordance concerning curricular content and its provision. Another scoping review [2] provided a thematic analysis of 41 publications on generative AI in medical education. The results revealed diverse potential applications such as self-directed teaching, simulation, and writing support. Nevertheless, the review also pointed on important matters (academic integrity, accuracy of information and possible negative effects on learning). Stamer et al. [22] reviewed 12 studies concerning training of healthcare professionals in communication and identified two main formats of AI-driven tools in this context: virtual reality simulation (VRS) with avatars and text or audio debriefing or evaluation. Performance of virtual patients was considered behind human performance in communication skills; therefore, motivation was shown as a main factor driving the use of AI-tools by learners. Chary et al. [20] systematically analyzed 14 studies reporting recent applications of natural language processing (NLP) in emergency medicine. The authors suggested that NLP could be useful to better track the progression of residents in emergency medicine. Levin et al. [26] systematically analyzed 19 studies that assessed ChatGPT's 3.5 performance in exams with multiple-choice questions. The authors identified heterogeneous design of studies with missing data, and potential overevaluation of ChatGPT's performance. Lucas et al. [13] provided a systematic analysis of 40 reports on the utilization of LLM in medical education. The study identified advantages of LLM such as personalized learning as well as concerns regarding data accuracy. Almost half of the papers included that review assessed performance of LLM in medical examinations. Other papers focused on future uses of LLM in medical education.

To our knowledge, the measurable effects of AI-based strategies on educational outcomes have not been systematically assessed [1–3, 24, 25]. The aim of this review was to perform a systematic analysis of the available literature regarding the measured effect of AI-powered strategies on educational outcomes in health professions education.

Methods

Data sources and searches

The literature search was conducted using the CINAHL Plus, EMBASE, Proquest, PubMed, Cochrane Library and Web of Science electronic databases to identify relevant articles in English and French until October 1st 2024. Published recommendations were followed to

create a search strategy [27, 28]. The search strategy is detailed in Table 1. The reference lists from the primary search were also searched manually to find additional eligible studies. For studies with multiple publications, only landmark trial reports were used as references [29]. Attempts were made to contact the corresponding

authors via email where publications were not available as full texts.

Study selection

This review considered studies analyzing a measured effect on the educational outcomes of AI-based tools/

Table 1 Search strategy

CINAHL Plus

1st Oct 2024

1974 – 2024

Find all my search terms

364 results

("medical student" OR "medical trainee" OR "health professional" OR "intern" OR "resident" OR "nurse") AND (ai OR a.i. OR artificial intelligence OR chat-GPT OR Chatbot OR chatterbot OR machine learning OR deep learning OR natural language OR neural network) AND (assist* OR coach* OR feedback OR suggest* OR guid* OR remote OR online) AND (assessment OR eval* OR metrics OR automatic OR electronic OR individual OR personalized OR target* OR motiv* OR autonomy* OR determin*) AND (dashboard OR display OR track* OR map*) AND (patient OR medical record OR clinical record OR health record OR notes OR outcome OR result OR performance OR competence)

EMBASE

1st Oct 2024

1974 - 2024

299 results

("medical student" OR "medical trainee" OR "health professional" OR "intern" OR "resident" OR "nurse") AND (ai OR a.i. OR artificial intelligence OR chat-GPT OR Chatbot OR chatterbot OR machine learning OR deep learning OR natural language OR neural network) AND (assist* OR coach* OR feedback OR suggest* OR guid* OR remote OR online) AND (assessment OR eval* OR metrics OR automatic OR electronic OR individual OR personalized OR target* OR motiv* OR autonomy* OR determin*) AND (dashboard OR display OR track* OR map*) AND (patient OR medical record OR clinical record OR health record OR notes OR outcome OR result OR performance OR competence)

Proquest

1st Oct 2024

1974 – 2024

Filters (scholarly journals or trade journals or other sources; Articles and revues; English and French)

505 results

("medical student" OR "medical trainee" OR "health professional" OR "intern" OR "resident" OR "nurse") AND (ai OR a.i. OR artificial intelligence OR chat-GPT OR Chatbot OR chatterbot OR machine learning OR deep learning OR natural language OR neural network) AND (assist* OR coach* OR feedback OR suggest* OR guid* OR remote OR online) AND (assessment OR eval* OR metrics OR automatic OR electronic OR individual OR personalized OR target* OR motiv* OR autonomy* OR determin*) AND (dashboard OR display OR track* OR map*) AND (patient OR medical record OR clinical record OR health record OR notes OR outcome OR result OR performance OR competence)

Medline Pubmed

1st Oct 2024

1974 – 2024

Search mode: all fields

3,858 results

("medical student" OR "medical trainee" OR "health professional" OR "intern" OR "resident" OR "nurse") AND (ai OR a.i. OR artificial intelligence OR chat-GPT OR Chatbot OR chatterbot OR machine learning OR deep learning OR natural language OR neural network) AND (assist* OR coach* OR feedback OR suggest* OR guid* OR remote OR online) AND (assessment OR eval* OR metrics OR automatic OR electronic OR individual OR personalized OR target* OR motiv* OR autonomy* OR determin*) AND (dashboard OR display OR track* OR map*) AND (patient OR medical record OR clinical record OR health record OR notes OR outcome OR result OR performance OR competence)

The Cochrane Library

1st Oct 2024

1974 – 2024

Reviews and Trials

161 results

(medical student OR medical trainee OR health professional OR intern OR resident OR nurse) AND (ai OR artificial intelligence OR chatGPT OR Chatbot OR machine learning OR deep learning OR natural language OR neural network) AND (assist OR coach OR feedback OR suggest OR guide OR remote OR online) AND (dashboard OR display OR track OR map) AND (patient OR medical record OR clinical record OR health record OR notes OR outcome OR result OR performance OR competence)

Web of sciences

1st Oct 2024

1974 – 2024

Core collections

795 results

(medical student OR medical trainee OR health professional OR intern OR resident OR nurse) AND (ai OR artificial intelligence OR chatGPT OR Chatbot OR machine learning OR deep learning OR natural language OR neural network) AND (assist OR coach OR feedback OR suggest OR guide OR remote OR online) AND (dashboard OR display OR track OR map) AND (patient OR medical record OR clinical record OR health record OR notes OR outcome OR result OR performance OR competence)

interventions in health professions assessment and/or training. Studies including health care professionals and undergraduate or postgraduate health care students or health care professionals were all considered eligible. Articles with no original data, studies not reporting educational outcomes, reviews, surveys, case reports, comments, letters, brief communications, book chapters, study protocols, preprints that had not undergone peer review, and studies published only as abstracts or conference proceedings were excluded. Furthermore, studies describing the development/validation of an AI-based technology, studies evaluating the performance/accuracy of AI-based technology/algorithms with no educational outcomes, and studies exclusively describing the experience or satisfaction of study participants were excluded. The inclusion and exclusion criteria for the present review are detailed in Table 2.

Data extraction

The titles and abstracts of all the citations identified by the literature search were reviewed independently by two investigators both experts in medical education (E.F. and H.H., who were blinded to the research question). Eligible articles were analyzed as full texts by the two investigators with 89% initial concordance (ICC 0.978 calculated using the IBM SPSS Statistics 30.0), which increased to 100% following discussion. Data collection was recorded on a shared MS Word document. All outcomes and

variables for which data were gathered are provided in the Tables and Supplementary material. The following data were extracted from the Cochrane Consumers and Communication Review Group’s data extraction template (<http://cccr.org/author-resources>): (a) author, year, geographical region, setting; (b) type of study: comparator group if applicable; (c) characteristics of IA-based strategy, characteristics of study participants, specialty; (d) inclusion criteria, exclusion criteria; (e) description of intervention, teaching modality, educational concept; (f) length of study, duration of follow-up; (g) outcomes, key conclusions; (h) and results of sensitivity analyses. Any missing information or disagreement in the data extracted by the two investigators was discussed, if necessary, with a third investigator (E.H-D. expert in medical education) and resolved.

Protocol and registration

Methods and inclusion criteria were defined in advance, specified in a protocol and registered in the OSF registries [30].

Quality assessment

All studies that fulfilled the inclusion criteria were assessed independently by E.F. and H.H. The appraisal tools used to evaluate the quality of the included studies are summarized in Supplementary Table 1. There are no appropriate frameworks for evaluating the quality of

Table 2 Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Published until May 13th 2024 in following databases: CINAHL Plus, EMBASE, Proquest, Pubmed, the Cochrane Library and Web of sciences or identified by manual search of references	Studies published after May 13th 2024 or personal communication not published in a peer-reviewed journal
Original study	No original data, reviews, comments, letters, study protocols, viewpoints, case reports or brief communications
Original study published in a peer-reviewed journal	Book chapters, preprints, or studies only published as abstracts or conference proceeding
Any kind of AI-based educational system with a description, such as machine learning (e.g., neural network to learn from data and perform tasks within a tutoring system)	AI-based educational system not specified
Health profession education that includes medical students, medical trainees, health profession trainees, health professionals, interns, residents, and nurses	Not health professions such as psychology
AI-based technology targeting training or assessment in health profession education	Studies focused solely on the development/validation of an AI-based technology or the evaluation of performance/accuracy of AI-based technology/algorithm
Focus on providing assessment of learning/behavior or performance, focus on monitoring/tracking learning progression or self-assessment of learning/training	Focus on providing training programs or scenarios without evaluation of educational outcomes
Address a measured effect on educational outcomes of AI-based tools/interventions in health profession education or training	Studies not reporting educational outcomes
Educational outcomes refer to measurable data that are collected by AI-based technology to support learning	Focus on data related to the satisfaction/experience with the use of AI-based technology or the satisfaction/experience with an AI-based educational strategy including qualitative studies

AI-powered interventions in health professions education. The existing TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis of Diagnosis) criteria [31] principally evaluate aspects regarding the development or evaluation of the performance of an AI-based prediction model and predicted outcomes. AI-based interventions need to be evaluated in several aspects: (i) external validation using separate data sets and specific algorithms; (ii) data accuracy and communication of uncertainty; (iii) approaches used for introducing decision support algorithms; (iv) sources of data, missing information and censoring [32]. We were unable to identify any appropriate frameworks to evaluate the quality of AI-powered interventions in health professions education. Therefore, the quality of interventions involving machine learning and AI was evaluated according to the recommendations of Bates et al. [32] (Supplementary Table 2). Randomized trials were assessed via the Consolidated Standards of Reporting Trials (CONSORT) guidelines (Supplementary Table 3) [33]. A checklist for artificial intelligence in medical imaging (CLAIM) was used for studies including AI in radiology [34] (Supplementary Table 4). The risk of bias was assessed using the Cochrane Collaboration's tools (RoB 2.0) for randomized trials [35] (Supplementary Table 5) and the appropriate version of risk of bias (ROBINS-1) for nonrandomized studies [36] (Supplementary Table 6). For observational studies, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [37] statement was used (Supplementary Table 7). Finally, assessment of certainty in the data records was performed using the GRADE approach [38] with five determinants of quality of the studies: (i) limitations of design and execution (risk of bias); (ii) inconsistency (heterogeneity); (iii) indirectness (components guiding the formulation of a research question); (iv) imprecision (number of items, number of participants, confidence intervals); and (v) publication bias. The GRADE framework enables one to reduce the level of certainty for each of the parameters and to make recommendations for practice: certainty very low, low, moderate and high (Supplementary Table 8). This approach is applied mainly in the evaluation of clinical studies [38]. The quality of the papers included in the present work was evaluated via the abovementioned tools independently by a second investigator (H.H.). Disagreements in data extraction were resolved by consensus, if necessary, after discussion with a third author (E.H-D.).

Data analysis

The primary outcome was a measured effect on the educational outcomes of AI in training and assessment in

health professions education, expressed as a performance score. The best practice in systematic reviews is to generate a meta-analysis of the outcomes to synthesize the results from numerous studies [29]. This requires studies to be sufficiently similar so that their outcomes can be pooled and compared [29]. Unfortunately, for this current review, the articles identified were heterogeneous in that a meta-analysis could not be usefully carried out. This review provides a synthesis of the included studies without a meta-analysis. The present analysis follows the PRISMA statement for systematic review articles [39] (Supplementary Table 9) and the Synthesis Without Meta-analysis (SWiM) guidelines [40] (Supplementary Table 10), given the absence of standards for reporting systematic reviews in health profession education. We also followed the STORIES statement (STructured approach to the Reporting In healthcare education of Evidence Synthesis) [41] (Supplementary Table 11) and BEME guidance [28].

To gain better insight into the educational impact, the analysis focused on educational theories and concepts used by different AI-powered interventions. In addition to outcome measures, the impact of AI-based educational interventions was evaluated according to the model proposed by Kirkpatrick [15] because of its capacity to assess learning effectiveness at different levels, including the workplace-based environment. As the intention was to examine the real impact of AI-powered educational interventions on the progression of learners, the focus was on the effects on educational outcomes rather than the attitudes or motivations of learners toward the use of AI-based tools. Therefore, this review explored levels 2, 3 and 4 in detail according to Kirkpatrick's model.

Herrington's authentic learning concept [42] was used to assess the authenticity of the AI-powered learning environments proposed by different studies. According to Herrington [43], an authentic learning environment may enhance the engagement of students and ameliorate learning outcomes. Herrington refers to the concept suggested by Brown et al. [44], indicating that purposeful learning has to be set to the environment within which it will be applied. Herrington indicates several elements constituting online situated learning, such as real-world context, activities with real-life relevance, rapid communication with experts, exploration of the topic from various points of view, collaborative activities and reflection on learning. On the basis of Herrington's concept, the following elements and their application by different studies were comprehensively searched: (i) authenticity in learning reflecting the modality in which the skills will be applied in real life; (ii) authentic activities having real practical relevance; (iii) mastery performance accessible to enhance learning; (iv) learning environment

exploring various perspectives; (v) collaborative activities to construct new knowledge; (vi) reflective practice; (vii) articulation of different learning activities; (viii) learning instructions and coaching; and (ix) authentic evaluation of learning.

Results

Literature search results

The search strategy yielded 6148 citations, 164 of which were duplicates, and 2 records were identified through manual search, generating a total of 142 records evaluated for eligibility (Fig. 1). Among these 142 studies, 130 were excluded for the following reasons: conference abstract ($n=3$), development or evaluation of an AI-based tool without educational outcomes ($n=41$), full text not available ($n=2$), letter to editor ($n=1$), no abstract available ($n=1$), no educational purpose ($n=2$), not an original study ($n=31$), preprints that had not undergone peer review ($n=6$), review ($n=12$), study protocol ($n=1$), and survey (perceptions/attitudes/knowledge about AI) ($n=30$) (see Supplementary Table 12 for details of the articles excluded from the review). This resulted in 12 studies being eligible for the final analysis. For two publications, which were not available as full texts, attempts to contact the corresponding authors via email were made but with no response.

Description of the studies

A total of 12 studies were included in the final analysis [45–56]. All studies were published between 2019 and 2024.

Three studies were conducted in Singapore [50, 51, 53], two in China [45, 55], two in the USA [54, 56], one in France [46], one in Germany [52], one in Italy [47], one in Japan [48] and one in Korea [49]. Four studies targeted nursing education [49–51, 53], one focused on interprofessional training [54], and the others focused on medical students, residents or physicians. The sample size ranged from 4 to 180 participants. All included studies were single-center studies. Ten studies included an experimental design. Of these, three [45, 48, 50] were randomized controlled trials, and seven had a quasi-experimental designs [47, 49, 51, 53–56]. Two studies were observational [46, 52].

Quality assessment results

Randomized trials were assessed using the CONSORT guidelines (Supplementary Table 3) and the Rob-2 tool [35] (Supplementary Table 5). Two randomized controlled trials had a high overall risk of bias [45, 50], and one had some concerns [48]. All quasi-experimental studies were rated to have a serious overall risk of bias (Supplementary Table 6). As detailed in Supplementary Table 2, the accuracy of the AI-based educational technologies was not evaluated, or information regarding the accuracy of the AI tool was not disclosed [45, 48–51, 53–55]. In one study [47], the accuracy of generating the diagnostic hypothesis via the AI tool was not assessed given the absence of an appropriate test. In the study by Zech et al. [56], the ground truth was established by consensus among three radiologists and did not include clinical data.

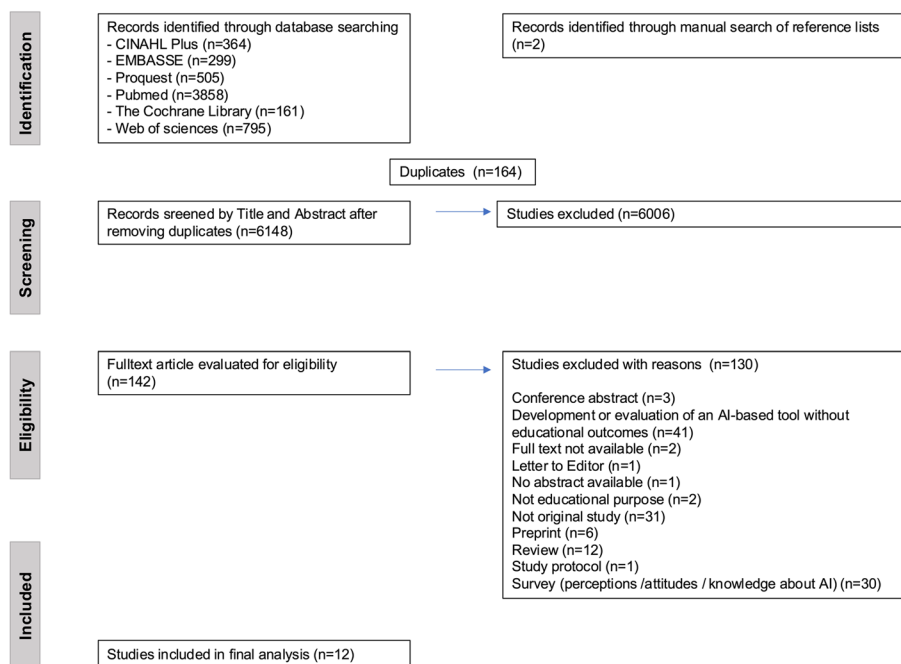


Fig. 1 Flowchart

Several studies may be biased in reporting their key findings [49, 51, 53–56]. The results of two studies were based on self-reported evaluations of educational outcomes [49, 53]. In one study [55], the students chose on a voluntary basis to participate in additional expert-led or expert- and AI-led sessions. In two studies [54, 56], repeated participation in the simulation scenario could influence the results. In the study by Zech et al. [56], participants reviewed the same set of cases both without and later with AI, overestimating potential improvements in clinical performance. The conclusions of the study were further complicated by the small size of each subgroup and variabilities in the level of training of the participants. A majority of the experimental studies were quasi-experimental and lacked appropriate controls. In both nonexperimental studies [46, 52], confounding variables were not controlled. In the study by Meetschen [52], the ground truth was not established by cross-sectional imaging but by the consensus of experts. A list of items according to the STROBE statement is presented in Supplementary Table 7.

None of the studies indicated what measures were undertaken to protect the personal data of learners and instructors, including the future use of the collected information. Copyright issues of commercially used AI-based tools were not reported [45, 46, 48, 52, 55, 56]. On the basis of the GRADE approach [57], the quality of evidence of experimental studies was considered low to very low. A certainty assessment of randomized controlled studies and quasi-experimental studies is presented in Supplementary Table 8. There was no statistical control for confounding variables. Owing to the heterogeneity of the identified articles and the overall risk of bias being judged as high, a meta-analysis was considered inappropriate. Consequently, a descriptive analysis was performed.

Objectives and outcomes of the studies

The aims and design of each study included in this analysis are detailed in Table 3. A description of the AI technology used by the included studies is provided in Supplementary Table 13. The main outcomes of the included studies are detailed in Table 4.

Experimental studies

A total of 10 experimental studies were included. The sample size ranged from 7 to 180 participants. The duration of the studies ranged from 20 min to 3 months. The following topics were covered by different studies: performance during endoscopic procedures [45]; medical interviewing [48]; clinical care [50]; interprofessional communication [50, 51, 54]; clinical reasoning [47]; confidence in medication administration [49]; image

interpretation [53, 56]; and ligature/suturing training [55].

AI-based tools have been used for various learning purposes: diagnosis and recognition [45, 55], medical interview assistance [48], medication administration [49], image interpretation [53, 56] and VRS [47, 50, 51, 54]. Three studies used large language models [47, 50, 51] to engage learners in VRS with avatars in human-like conversations. The results of a randomized controlled study that included 65 nursing students by Liaw et al. [50] revealed that 2 h of AI-powered VRS improved the sepsis care knowledge of nursing students; however, a human-controlled group significantly improved self-efficacy in interprofessional communication. In a quasi-experimental study that included 32 students, Liaw et al. reported promising results of AI-powered VRS in enhancing the knowledge of students regarding skills in interprofessional communication [51]. Comparable results were shown by Shorey et al. [53], suggesting potential benefits of AI-based VRS in improving the communication self-efficacy of students. However, the lower scores for clinical communication performance indicated some concerns regarding the authenticity of virtual patient simulation (VPS). All three studies suggested that future developments are needed to increase the benefits of AI-powered avatars in teaching communication skills. Truong et al. [54] proposed virtual reality (VR)-based simulation training in the management of operating room fires. More than 50% of the study participants (97/180) successfully passed after the 4th or 5th tentative (VR with help by AI), with only 24% passing (43/180) without AI. The authors demonstrated the feasibility and effectiveness of the VR approach powered by AI for teaching these skills. Furlan et al. [47] proposed VPS for clinical diagnostic reasoning. The authors provided outcomes of a knowledge test administered shortly after 2 h of VPS to a group of 15 undergraduate medical students. The performance of two students worsened after simulation, and that of one student did not change. The authors suggested that the combination of intelligent tutoring system (ITS) and VPS AI-powered strategies may represent a training tool in diagnostic reasoning in situations in which students have limited possibilities to attend clinical wards. Kim et al. [49] developed a learning chatbot for medication administration that enhances the self-directed learning skills of students. The chatbot was designed for new nurses without clinical experience. A 2-week trial, which included 17 participants, revealed improvements in the confidence of the nurses in their medication administration knowledge. Kanazawa et al. [48] reported that an AI-powered tool for conducting medical interviews improved diagnostic skills and decreased the consultation times of participants ($n=20$). An & Wang [45] and Yang & Shulruf [55]

Table 3 Aims and design of the included studies

Author/Year	Aim	Design	Study duration
Randomized controlled studies			
An et Wang 2023 [45]	To evaluate the impact of AI-powered diagnostic and detection system in training of gastroscopy on students' performance	Single center	3 months
Kanazawa 2023 [48]	To assess the efficacy (i.e., amelioration of diagnostic precision) of an AI-powered medical interview-assistance system	Single center, open-label, crossover trial	20 min
Liaw 2023 [50]	To assess the efficacy of an AI-controlled physician compared to a human-controlled physician to train nurse students in the sepsis care and in interprofessional communication	Single center trial (pretest–posttest design)	2 h (2 simulation scenarios)
Quasiexperimental studies			
Furlan 2021 [47]	To construct VPS based on natural language associated with ITS for diagnostic reasoning	Single center, one-group pretest–posttest design	2 h
Kim 2024 [49]	To construct a chatbot promoting the self-directed learning capacities of new nursing students on medication administration	Single center one group pretest–posttest design	2 weeks
Liaw 2023 [51]	To evaluate students' competencies in communication with an AI-based physician in a VRS context	Single center mixed-method, one-group pretest–posttest design	2 h
Shorey 2023 [53]	To evaluate the application of VPS in preparing nurse students for communication with patients and health care providers	Single center, pretest and posttest design, single group	2-years
Truong 2022 [54]	Ability of health care professionals to solve an AI-assisted VRS operating room fire scenario	Single center pre/posttest study	2 weeks
Yang et Shulruf 2019 [55]	To evaluate whether an AI-based system tutoring course improves suturing/ligature skill acquisition by medical interns	Single center comparative study	10 weeks
Zech 2024 [56]	To evaluate the effect of an openly available AI model on physicians' capacity to interpret upper extremity pediatric radiographs	Single center pre–posttest design	4 weeks
Observational studies			
Chassagnon 2023 [46]	To assess the impact a DL-based system on interpretation of chest X-rays (pneumothorax, pleural effusion, alveolar syndrome, pulmonary nodule, and mass in mediastinum) by residents in radiology	Single center retrospective case–control study	3 months
Meetschen 2024 [52]	To assess the efficacy of an AI deep neural network-algorithm to detect fractures by residents in radiology	Single center retrospective descriptive study	Not specified

AI artificial intelligence, DL deep learning, ITS intelligent tutoring system, VPS virtual patient simulation, VRS virtual reality simulation

reported the potential value of AI-led systems in learning technical skills. Zech et al. [56] reported potential implications of AI-based imaging interpretation assistants in improving the accuracy of residents ($n=5$) in detecting fractures compared with attending subspecialized training.

Non-experimental studies

Both included non-experimental studies [46, 52] had retrospective designs and small numbers of participants. Chassagnon et al. [46] reported that AI improves the performance of radiology residents ($n=8$) in the

interpretation of pulmonary radiographs; however, AI cannot be used alone as a teaching strategy. Similarly, Meetschen et al. [52] reported that AI support improved the ability of radiology residents to identify fractures ($n=4$), indicating its potential application as a training tool.

Teaching modalities and theoretical concepts used by studies

The theoretical concepts and teaching modalities used by the included studies are presented in Table 5. Simulation was the most common teaching modality reported by 8 studies [45, 47, 48, 50, 51, 53–55]. Of these, six studies

Table 4 Main outcomes of the included studies

Author/Year	Participants	Outcomes
Randomized controlled studies		
An et Wang 2023 [45]	32 graduate students in gastroenterology: Intervention (conventional teaching of gastroscopic procedure + AI) (N=16) Comparator = conventional teaching (N=16)	Final performance of the gastroscopic procedure AI group compared to non-AI group: - higher performance score, success rate, and detection rate of lesions ($P < 0.05$) - lower mean time and reduced patient pain scores ($P < 0.05$)
Kanazawa 2023 [48]	Two trials with a crossover for a total of 200 differential diagnoses: 20 first-year resident physicians (10 per trial): AI group (N=5) Non-AI group (N=5)	192 differential diagnoses of participants were analyzed for a correct answer rate by three experienced physicians AI group compared to non-AI group: - higher rate of correct diagnosis ($P=0.02$) - lower consultation time ($P=0.04$)
Liaw 2023 [50]	Nurse participants (N= 64) randomly assigned to 2 groups: AI-group or instructor-controlled group	AI- versus instructor-led groups: higher sepsis knowledge score ($P=0.009$); no differences in performances in sepsis care ($P=0.39$) and interprofessional communication ($P=0.21$)
Quasixperimental studies		
Furlan 2021 [47]	Undergraduate 5 th year medical students (N=15) underwent two identical tests (before and after simulation) composed of 22 multiple-choice questions.	Presimulation vs. postsimulation performance: increase in mean test score ($P < 0.001$) Two students' performances worsened, one student without change.
Kim 2024 [49]	Novice nurses (N=17), self-directed learning. Evaluation: online survey (before and after intervention)	Increased confidence of the participants in medication administration knowledge ($P < 0.001$).
Liaw 2023 [51]	Final year nursing students (N=32)	Significant improvements in knowledge and self-efficacy in interprofessional communication
Shorey 2023 [53]	Undergraduate nursing students (N= 93): VP trainings with 4 scenarios. Comparator group: historical control group of students.	AI-based group vs. historical group: - improvements in students' learning attitudes toward communication skills - lower scores in pediatric, obstetric, and medical practicum on clinical communication.
Truong 2022 [54]	Operating room health care professionals (N= 180) Five VRS trials (2 final trials with AI)	Improvements in knowledge and performance in OR fire management: N= 8 (4.4%) success in the first simulation trial, N=43 (23.9%) success on the 3rd attempt (VR only); N= 97 (53.9%) success on the 4–5th tentative attempt (VR + AI). Expert-led and expert-led + AI groups vs. traditional group: better block OSCE scores
Yang et Shulruf 2019 [55]	3 groups of surgery residents: Group 1 (regular training, N=25) Group 2 (traditional training + instructor-led, N=24) Group 3 (traditional training + instructor-led + AI, N=23)	
Zech 2024 [56]	Trained radiologists (N=2), radiology residents (N=3), pediatric residents (N=2): interpretation of 240 cases without and with AI (3–4 weeks later)	Improvements in accuracy in identifying upper extremity fractures of residents ($P < 0.001$). No significant effect on the accuracy of subspecialized musculoskeletal radiologists. Reduction in interpretation time to 38.9 s with AI vs. 52.1 s without AI ($P=0.030$).

Table 4 (continued)

Author/Year	Participants	Outcomes
Observational studies Chassagnon 2023 [46]	Radiology residents (N=8) Step 1: interpretation of 150 CXRs Step 2: interpretation of 200 CXRs Group 1: AI-based system 2 nd reader vs. Group 2: control Step 3: interpretation of 150 CXRs	Group 1 vs. Group 2: At intervention: increased sensitivity (53% vs. 43%), specificity (94% vs. 90%), and accuracy (86% vs. 81%) ($P < 0.001$) After intervention: no differences in sensitivity (44% vs. 46%), specificity (90% vs. 90%), or accuracy (80% vs. 80%).
Meetschen 2024 [52]	Radiology residents (N=4) evaluated 100 radiographs with at least 1 fracture and 100 radiographs without a fracture with and without AI.	Improvements in sensitivity for detecting fractures with AI assistance (58% vs. 77%; $P < 0.001$) No significant amelioration in specificity. Reduction in interpretation time with AI by 2.6 s ($P = 0.0156$)

AI artificial intelligence, CXR chest X ray, DL deep learning, ITS intelligent tutoring system, MLP natural language processing, VP virtual patient, VRS virtual reality simulation

Table 5 Theoretical concepts and teaching modalities used by the studies

Author/Year	Teaching modality	Learning theory or concept
Randomized controlled studies		
An et Wang 2023 [45]	Instructor-led on-site simulation sessions	Not reported
Kanazawa 2023 [48]	On-site simulation sessions: mock medical interviews with simulated patients	Not reported
Liaw 2023 [50]	AI-powered group: individually with support via Zoom chat by the research team. Human-controlled group: 4–6 instructors and facilitators.	Experiential learning and social constructivism Communication strategies [58]
Quasiexperimental studies		
Furlan 2021 [47]	VPS combined with NLP and intelligent tutoring system to train the analytical thinking of students	Not reported
Kim 2024 [49]	A chatbot through smartphone-based learning (self-directed learning) Form: providing knowledge, quizzes, and encouraging messages	Not reported
Liaw 2023 [51]	Remote AI-enabled VRS while communicating with instructors (Zoom chat) regarding any questions	Kiili's experiential gaming model [60] incorporating feedback Communication strategies adapted from Ross et al. [58]
Shorey 2023 [53]	VRS: VP scenario conducted by a research assistant over 1-hour sessions (25 students per session) at the university's computer laboratory. Each student had access to the VP scenarios. Feedback on communication at the end of the session.	Bandura's self-efficacy theory [59] and Herrington's authentic learning concept [42]
Truong 2022 [54]	VRS with one participant during the scenario	Not reported
Yang et Shulruf 2019 [55]	Simulation led by experts	Not reported
Zech 2024 [56]	Not reported	Not reported
Observational studies		
Chassagnon 2023 [46]	Not reported	Not reported
Meetschen 2024 [52]	Not reported	Not reported

AI artificial intelligence, NLP natural language processing, VPS virtual patient simulation, VR virtual reality, VRS virtual reality simulation

integrated VRS [47, 50, 51, 53, 54]. One study proposed self-directed learning [49], and in three studies, the teaching modality was not specified [46, 52, 56]. Three studies [50, 51, 53] reported the use of learning theories for the development of AI-powered educational strategies. Experiential learning was leveraged in two studies in the context of AI-powered VRS [50, 51]. Both studies taught communication strategies fostering the capacities of learners to work in teams and leadership [58]. One study [53] applied Bandura's self-efficacy theory [59] and Herrington's authentic learning concept [42] to promote the communication skills of nursing students via AI-powered VPSs. Several studies [47, 49–51] reported that feedback activities and/or prompts were incorporated into educational interventions. Feedback activities and prompts were variable, such as support via Zoom chat by the instructors [50, 51], chatbots delivering encouraging messages [49], trained research assistants [53] or ITSs [47]. No detailed overviews of the activities used during the educational intervention were found in any of the articles. No information was provided on the impact of feedback activities on learning outcomes.

Authenticity of the learning environment based on Herrington's concept

Shorey et al. [53] used Herrington's authentic learning concept [42] in the context of the development of AI-powered VPSs to teach nursing students communicating skills. The remaining studies reported neither a theoretical framework nor an analysis of the authenticity of the learning environment [45–52, 54–56].

We conducted a detailed analysis of the authenticity of the AI-powered learning environment according to Herrington's concept [42]. The results for all twelve studies are provided in Supplementary Table 14. The majority of the studies did not consider all nine items proposed by Herrington. Two underrepresented items were reflective practices, which were incorporated by four studies [47, 50, 51, 53], and authentic evaluations of learning, which were considered by four following studies [45, 47, 53, 55].

Educational impact according to the model proposed by Kirkpatrick

The model proposed by Kirkpatrick was applied to obtain better insight into the educational impact of the

included studies [15]. Learning outcomes measured by the included studies are presented in Table 5. Among the specific variables measured were correct diagnosis [48], diagnostic accuracy [46, 52, 56], communication skills [50, 53], level of self-confidence [49], self-efficacy [51] or learning aptitudes [53], performance in OSCE [55], performance in specific care [50], procedural skills [45] and specific knowledge scores [47, 50, 51, 54]. All experimental studies [45, 47–51, 53–56] and nonexperimental studies [46, 52] attained Kirkpatrick's level 2. None of the studies reached levels 3 or 4. Level 2 corresponds to the evaluation of learning. It is the assessment of acquired knowledge after educational intervention [15]. As detailed in Supplementary Table 15, all studies, except for five [45, 46, 48, 52, 55], applied the assessments before and after the educational experience; however, scoring instruments were not provided.

Discussion

General interpretation of the findings

The results of the analysis of the twelve studies indicate that the current evidence regarding the educational outcomes of AI-powered interventions in health professions education is poor. The overall impact of AI-based educational tools is supported by the results of single-center studies with small numbers of participants and a limited duration. Therefore, the findings of the present review cannot fully answer the research question regarding measurable outcomes of AI-based educational interventions in health professions education. Nevertheless, a unique feature of the current review is identification and in-depth analysis of conceptual frameworks of learning theories applied within AI-based educational interventions. Learning outcomes cannot be meaningfully measured without the analysis of other components related to AI interventions, such as learning theories, learning tools, instructional design, and interactions between the learner and AI-based tools. Indeed, as the learning strategies are potentially different, learners may face some challenges when addressing tasks in AI-based learning environments [61]. In the present review the impact of AI-based educational interventions was evaluated according to the Kirkpatrick's model [15] given its capability to assess learning effectiveness at different levels. We focused on the levels 2, 3 and 4 to explore effects on educational outcomes rather than assessing the reactions of learners. Based on the Herrington's authentic learning concept [42] the present review comprehensively assessed the authenticity of the AI-powered learning environments proposed by different studies. The use of the concept of authentic learning by different studies was inconsistent. The lack of learning

content in most AI-based educational tools is also emphasized, asking whether a new approach should be considered to adopt learning theories as a basis for the construction of AI-powered educational tools. The present work also highlights that there is no straightforward guide for evaluating the quality of research in AI-based education and suggests a series of criteria that should be considered.

Application of learning theories and principles

Regarding conceptual frameworks applied within AI-based educational interventions, one study [53] used the principles of social learning theory [62] and Bandura's self-efficacy [59] in the context of AI-powered VRS. When applying this principle, students observe their instructor doing clinical tasks such as medical examination and consultation [63]. In the virtual environment, this can be represented, for example, by an avatar playing the role of a senior physician. However, the use of avatars in the studies included in the present review was considered less authentic (such as voice, expressiveness or communication) [50, 51] and may be less appropriate than the use of real simulated patients. This may indicate that future developments should focus on blended learning, where AI-powered educational strategies are followed by in-person simulation with real human-to-human interactions.

Herrington's concept was reported as a theoretical framework in the development of AI-based tools by one study [53] with the aim of enhancing the communication skills of learners. The authors used the principle of situated learning while applying social learning theory [62] and Bandura's self-efficacy [59]. However, in the study by Shorey et al. [53], each student executed learning tasks on the computer once and without interactions with peers. As suggested by Bandura, self-efficacy influences the preferences of learners and their perseverance in learning activities. This approach might be suitable for self-learning activities proposed by some AI-based educational tools. One such example is the study by Kim et al. [49]. In that study, a chatbot provided guidance and new knowledge through quizzes, prompts and encouraging messages.

Experiential learning is another example of self-directed learning, and a key principal of constructivism [64]. It was used by two studies analyzed in this review in the context of AI-powered VRS [50, 51]. According to Kolb [65], new knowledge is acquired through personal experience and reflective activities. Educators help learners enhance their reflection in the construction of a new experience [66]. Both studies [50, 51] mentioned interactions with educators via Zoom; however, there was no detailed information regarding the character of these

interactions. Furthermore, they did not report whether the learning objectives of the students were taken into consideration. Indeed, new knowledge is acquired more easily if the learning strategy considers the aims of learners, showing them an immediate use of new knowledge in real-life settings [67].

It has to be mentioned here that AI and VR are two different technologies that have different purposes. VR refers to the creation of immersive virtual environments that simulate the real world and it is often used in health professions education to enhance learner engagement and create immersive experiences, such as a robotic surgery training [68]. On the contrary, AI relates to the development of computer systems that can perform tasks that typically require human intelligence (e.g., natural language processing or decision-making). AI and VR can be combined to create more elaborated educational strategies. For example, AI-powered chatbots can be used within VR environments to provide learners with personalized recommendations and support [50, 51].

Analysis of learning outcomes

None of the studies included in the current work provided learning objectives or descriptions of the competencies to be achieved. Learning objectives play a principal role, as both learners and their supervisors should have identical understandings of the proposed objectives [69]. The learner should have the ability to think critically about his own learning experience and should be capable of confronting it with the application of different theories according to specific learning goals. As proposed by Linn & Miller [70], learning products rather than the learning process should be the focus of predefined learning outcomes. In addition, it is important to be familiar with the technology used in the context of the educational process [71]. Furthermore, feedback activities and prompts promote self-directed learning, enhance the motivation of students and deepen their learning. Feedback contributes to the amelioration of learning outcomes and develops the responsibility of learners [72]. Several studies included in this review [47, 49–51, 53] incorporated feedback activities into educational interventions. However, the authors did not report feedback activities in detail, and it is difficult to analyze their impact on learning outcomes. The learning outcomes reported by the studies analyzed in this review correspond to level 2 (learning) according to Kirkpatrick [15]. Kirkpatrick's second level is relevant for some training achievements, such as technical skills or quantifiable knowledge. However, it cannot evaluate more complex tasks, such as the behavior of learners in the workplace [15]. In addition, challenges arise if the technology is

poorly designed, which may compromise the measurement and strategy of data analysis.

As indicated in Supplementary Table 2, none of the studies reported whether the learners and their instructors were informed about the accuracy of predictions of AI-based tools. This is important because it impacts learning outcomes and their assessment. Similarly, none of the studies reported whether the learners and their instructors were familiar with AI-based technologies. This aspect has already been highlighted in the literature, and previous reviews have expressed the need for improvements in the AI literacy of users, as well as their ability to critically evaluate the accuracy of AI-based tools [2, 13, 20–22, 24]. Given the growing place of AI-based technology in health care, new skills that health professionals should master have started to emerge [73]. Jacobs et al. [73] proposed that medical graduates should be capable of demonstrating an understanding of the principles of health informatics and how to apply them, using AI-powered platforms and develop interdisciplinary collaborative practices, recognizing the limits of AI-based technologies, including data management and data privacy, and considering the benefits of data use against potential risks. This suggests some avenues for future exploration in this area.

Transparency in reporting the use of AI-based technologies

We identified several gaps in the transparent reporting of AI-based tools by the included articles, such as a lack of information on the accuracy of AI tools and the potential risk of errors. There was insufficient information on training datasets and copyright issues. Note that AI algorithms may challenge classical bases of teaching and learning, as it may be difficult to define learning outcomes and evaluate student performance in the case of AI tool errors. Like in clinical practice [74], the risk of biased learning from biased training data could compromise learning efficiency for both students and their instructors [75].

Furthermore, the studies did not report how personal data were handled, how learning outcomes were evaluated or where these pieces of information were stocked. There was no information on the future use of these data. The privacy of learners' data and instructor data implies respect for regulatory obligations [76]. There were no data in the studies about the measures that were undertaken to ensure that personal data were protected. This is an important point and should be considered when designing AI-based educational strategies.

The present work indicates that there is no straightforward guide for evaluating the quality of research in

AI-based education and suggests a series of criteria that should be considered:

- Algorithm used and its validation.
- Training datasets and standards used.
- Accuracy of the AI tool.
- Copyright issues.
- Handling of personal data (learning analytics, video recordings, etc.)
- Archiving and use of generated data.
- Definition of learning outcomes.
- Description of educational strategy (learning theory, instructional design, duration of learning activity, etc.)
- Evaluation of learning outcomes.
- Evaluation of teaching efficacy.
- Feedback modalities: students, instructors, developers.
- Program efficiency and its evaluation.

Key challenges and suggestions for the future

Several key challenges highlighted by the studies in this review, as well as suggestions for the future, are illustrated in Fig. 2. There is a need for more robust studies and more rigorous measures of learning outcomes to evaluate the effectiveness of AI-powered educational strategies. The accuracy of AI tools and information on potential errors in the system need to be communicated to learners and their teachers. The absence of clearly defined learning objectives in the reviewed studies is emphasized. It is desirable to apply learning theories when designing

AI-based educational tools to engage learners. Learning theories can help designers develop more effective AI-based educational tools and choose specific instructional strategies [63]. The use of identified theoretical models, such as Kirkpatrick’s levels [15] and Herrington’s framework [43] or active learning design principles [77], is suggested to guide the design of AI-based educational tools. The measurable achievements (e.g., skills) that a learner should perform [78, 79] must be defined to enable learners to accomplish their professional roles.

Furthermore, there are some concerns related to physical and emotional interactions due to the virtual environment. Artificial intelligence-powered educational strategies can increase the core knowledge scores of learners; however, human-led groups show greater self-efficacy in interprofessional communication [51] and diagnostic reasoning [47]. The combination of VRS with blended learning environments or further improvements in the design of AI tools could facilitate a more socially powered AI. Such an example could use emotional scales in training data for virtual agents [51]. Indeed, social interactions improve attitudes toward interprofessional collaboration. According to Bandura [80], learning occurs through observation and interaction with others in a social-environmental context. A recent scoping review on this topic [22] (including 12 studies) focused on undergraduate health care education and reported that the motivation of learners plays a crucial role in the use of these technologies because of the limits regarding the authenticity of VRS in illustrating real situations.

Strategies that favor learning effectiveness are lacking. AI-based pedagogical agents (such as 3D doctors) that

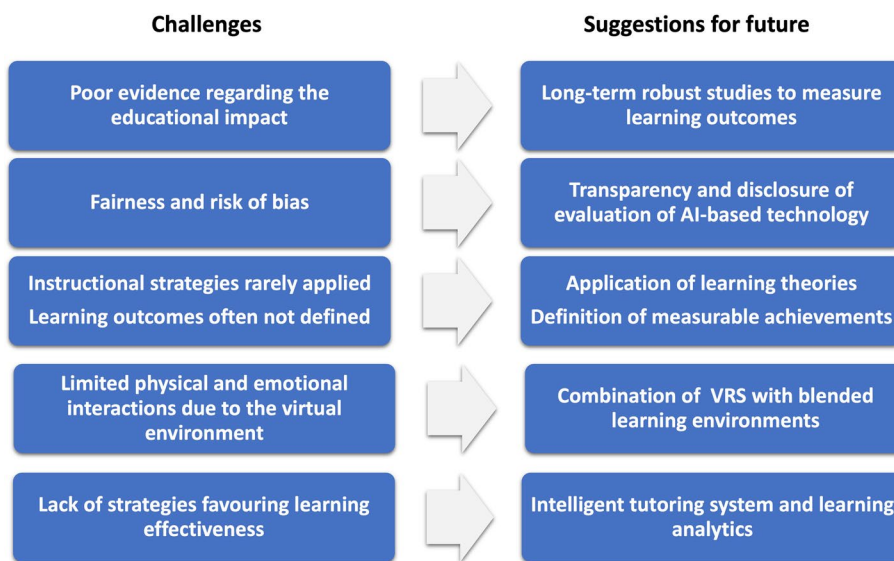


Fig. 2 Key challenges of AI-powered educational strategies in health professions education highlighted by studies and suggestions for the future

engage students and motivate them during self-learning, for example, by providing feedback and prompts, could facilitate knowledge construction and the development of the reasoning process of learners [51]. In addition, ITSs using AI can provide personalized guidance for learners [81]. AI-based learning analytics can be implemented in VRS tools, provide feedback to students and inform teachers about the learning outcomes of their students. Moreover, this has the potential to support sustainable education [81]. Notably, only one study [47] included in the present review incorporated ITSs within the AI-based learning tool, but the evaluation of learning outcomes and competencies to be achieved was not defined.

Finally, no study included in the current review reported financial issues and cost-effectiveness. Interestingly, a recent review [82] indicated that among the principal reasons for use of AI in medical education is its capacity to decrease costs. Therefore, another aspect to be considered by future studies is the cost-effectiveness and scalability of AI-powered educational tools.

This review has identified new educational perspectives that give health profession educators concrete examples of how to apply AI in their teaching activities. In the era of AI, medical teachers must also consider the transformative aspects of health care frameworks, changing demands on the quality of provided care and other elements, such as the benefits and potential limits of AI in terms of learning. All these aspects are translated into the evolution of teaching modalities; therefore, mastering learning theories by clinical teachers becomes a desirable necessity.

Medical schools and AI professionals play a key role in preparing medical students for the new competencies required as a result of AI in health care [83]. In the context of AI-powered clinical practice, this can also represent the creation of novel competencies to respond to emerging health care needs [84]. The American College of Physicians has recently published a position paper analyzing the role of AI in health care, including its potential advantages and concerns [85]. Among the current challenges of implementing AI-based education are the lack of standardization of curricula [82, 86]. Interestingly, some universities such as the National University of Singapore offer compulsory undergraduate program in bioinformatics and AI in medicine [87], which may explain the fact that several studies included in this review are from the research groups from Singapore [50, 51, 53].

Significance and transferability of the findings

All studies included in this review focused on a specific topic and were monocentric. In addition, their overall quality was low, limiting the transferability of the results to other educational settings. Moreover, a lack of expert

educators in AI may limit the transferability to other institutions [82] and lead to inequities in access to AI-powered education. Finally, as mentioned earlier in this work, AI-based technologies have the potential to support sustainable education [81]; however, it is important to consider an obsolescence index that applies to AI tools given the rapid evolution of medical knowledge [88].

One should consider the significance of AI-based educational tools for society and not for individuals. Indeed, AI can improve collaborative practices and prepare future health professionals for evolving health care systems [89]. To enhance collaborative activities among learners, AI-based tool can be designed for the groups of learners to solve more complicated clinical situations. In parallel the educators should find an equilibrium between AI-led and human-led educational strategies to prevent an increased dependence of learners on AI-tools [90]. Educational design research may bring new insights and enhance collaboration between stakeholders to develop both new theoretical understandings and practical strategies [91]. Another point to be mentioned is copyright issues regarding both the utilization of AI-based tools and the reporting of data in the scientific literature [18].

Strengths and limitations

This is the first study to analyze the impact of AI-based educational interventions in health profession education via a robust search strategy. Although only studies in the English language were included in this analysis, the search strategy did not include language filters, and no study was excluded because it was not published in English.

The results are based on the search strategy for the pre-defined keywords and databases used. The selection of databases was based on their free availability or access via university libraries. The search strategy followed the PICO model [92]. Despite recent data not supporting the recommendation to search for outcomes [93], we considered outcomes as an important part of the search strategy, as they constituted the research question. In addition, the GRADE recommendations [38] allowed assessment of the level of certainty of the studies regarding a risk of bias, precision, consistency, and directness.

We focused on the characterization and analysis of the learning theories and principles used by the authors of the studies in the development of AI-based teaching strategies. These were considered illustrative and concise but are not meant to be exhaustive. The objective was to provide concrete examples for health profession educators and developers of AI-based tools while providing new perspectives for future evolution. In parallel, several main challenges are indicated with suggestions for improvements.

The studies had a heterogeneous design; therefore, a meta-analysis was not possible. The heterogeneity of the included studies also limits the generalizability of the findings. Moreover, the potential biases introduced by small sample sizes and single-center designs have to be acknowledged here as these factors could significantly influence the validity of the results. There are no appropriate reporting standards or guidelines to evaluate AI-powered educational strategies, including the evaluation of the risk of bias [94]. Thus, when available, guidelines for AI-powered studies were used; otherwise, existing standards for non-AI-powered studies were employed in this work. In addition, a low comparability between the different assessment instruments needs to be mentioned here. All the studies were of short duration; consequently, it is not possible to evaluate the sustainability of these educational strategies. Significant concerns about privacy were noted. None of the studies indicated what measures were undertaken to protect the personal data of participants. Copyright issues of commercially used AI-based tools were not disclosed.

Conclusions

The results of the analysis indicate that the current evidence regarding the educational outcomes of AI-powered interventions in health professions education is poor. Further studies with a rigorous methodological approach are needed. It is desirable for educators and developers to integrate AI into blended learning environments that balance technological capabilities with human interaction. The present work also highlights that there is no straightforward guide for evaluating the quality of research in AI-based education and suggests a series of criteria that should be considered.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-025-06719-5>.

Supplementary Material 1.

Authors' contributions

E.F., E.H-D. were involved in the conception and design of the study. E.F., H.H. and E.H-D. analyzed and interpreted the data. E.F. drafted the paper. E.F. and E.H-D. revised it critically for intellectual content. E.F., H.H. and E.H-D. approved the version to be published and agree to be accountable for all aspects of the work.

Funding

The present work had no funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculté de médecine, maïeutique et métiers de la santé, Université de Lorraine, Nancy, France. ²Centre universitaire d'enseignement par simulation (CUESIM), Hôpital virtuel de Lorraine, Université de Lorraine, Nancy, France. ³Institut national de la santé et de la recherche médicale (Inserm), Unité mixte de recherche (UMR) U1116 – Défaillance cardiovasculaire aiguë et chronique (DCAC), Université de Lorraine, Nancy, France. ⁴Centre for Medical Education, University of Dundee, Dundee, UK. ⁵Centre Universitaire d'Enseignement par Simulation - CUESim Hôpital Virtuel de Lorraine – HVL, Faculté de Médecine, Maïeutique et Métiers de la Santé, 9, Avenue de la Forêt de Haye, Vandœuvre-lès-Nancy 54505, France.

Received: 31 October 2024 Accepted: 20 January 2025

Published online: 27 January 2025

References

- Karabacak M, Ozkara BB, Margetis K, Wintermark M, Bisdas S. The Advent of Generative Language models in Medical Education. *JMIR Med Educ.* 2023;9:e48163.
- Preiksaitis C, Rose C. Opportunities, challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: scoping review. *JMIR Med Educ.* 2023;9:e48785.
- Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large Language models in Medical Education: applications and implications. *JMIR Med Educ.* 14 août. 2023;9:e50945.
- Abbott KL, George BC, Sandhu G, Harbaugh CM, Gauger PG, Erkin O, et al. Natural Language Processing to Estimate Clinical Competency Committee Ratings. 2021;78(6):2046–51.
- Johnsson V, Søndergaard MB, Kulasegaram K, Sundberg K, Tibblad E, Herling L, et al. Validity evidence supporting clinical skills assessment by artificial intelligence compared with trained clinician raters. *Med Educ.* 2024;58(1):105–17.
- Winkler-Schwartz A, Yilmaz R, Mirchi N, Bissonnette V, Ledwos N, Siyar S, et al. Machine Learning Identification of Surgical and Operative factors Associated with Surgical expertise in virtual reality Simulation. *JAMA Netw Open.* 2019;2(8):e198363–198363.
- Masters K. Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158. *Med Teach.* 2023;45(6):574–84.
- Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. 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.* 8 févr. 2023;9:e45312.
- Roos J, Kasapovic A, Jansen T, Kaczmarczyk R. Artificial intelligence in medical education: comparative analysis of ChatGPT, Bing, and medical students in Germany. *JMIR Med Educ.* 4 sept. 2023;9:e46482.
- Knopp MI, Warm EJ, Weber D, Kelleher M, Kinnear B, Schumacher DJ, et al. AI-Enabled Medical Education: threads of Change, Promising futures, and Risky realities Across four potential future worlds. *JMIR Med Educ.* 2023;9:e50373.
- Ahsan MM, Luna SA, Siddique Z. Machine-learning-based Disease diagnosis: a Comprehensive Review. *Healthcare.* 2022;10(3):541.
- Alowais SA, Alghamdi SS, Alsuhbany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ.* 2023;23(1):689.
- Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. *Med Educ.*

- 19 avr 2024 [cité 4 juill 2024];n/a(n/a). Disponible sur: <https://doi.org/10.1111/medu.15402>.
14. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication [published erratum appears in *Acad Med*. 1992;67(4):287]. *Acad Med*. 1990;65(10). https://journals.lww.com/academicmedicine/fulltext/1990/10000/a_cognitive_perspective_on_medical_expertise_.1.aspx.
 15. Kirkpatrick DL. Techniques for evaluation Training Programs. *J Am Soc Train Dir*. 1959;13:21–6.
 16. Maicher KR, Stiff A, Scholl M, White M, Fosler-Lussier E, Schuler W, et al. Artificial intelligence in virtual standardized patients: combining natural language understanding and rule based dialogue management to improve conversational fidelity. *Med Teach* 4 mars. 2023;45(3):279–85.
 17. Ellaway RH, Tolsgaard M. Artificial scholarship: LLMs in health professions education research. *Adv Health Sci Educ* 1 août. 2023;28(3):659–64.
 18. Abd-alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large Language models in Medical Education: opportunities, challenges, and future directions. *JMIR Med Educ* 2023;9:e48291. <https://doi.org/10.2196/48291>.
 19. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and Generative Artificial Intelligence for Medical Education: potential impact and opportunity. *Acad Med*. 2024;99(1). Disponible sur: https://journals.lww.com/academicmedicine/fulltext/2024/01000/chatgpt_and_generative_artificial_intelligence_for.11.aspx.
 20. Chary M, Parikh S, Manini AF, Boyer EW, Radeos MA. Review of Natural Language Processing in Medical Education. *Western J Emerg Med*. 2019;20(1):78–86.
 21. Preiksaitis C, Ashenburg N, Bunney G, Chu A, Kabeer R, Riley F, et al. The role of large Language models in transforming Emergency Medicine: scoping review. *JMIR Med Inf* 10 mai. 2024;12:e53787.
 22. Stamer T, Steinhäuser J, Flägel K. Artificial Intelligence Supporting the Training of Communication Skills in the education of Health Care professions: scoping review. *J Med Internet Res*. 2023;25:e43311.
 23. Tolsgaard MG, Boscardin CK, Park YS, Cuddy MM, Sebok-Syer SS. The role of data science and machine learning in Health professions Education: practical applications, theoretical contributions, and epistemic beliefs. *Adv Health Sci Educ Theory Pract* déc. 2020;25(5):1057–86.
 24. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide 84. *Med Teach* 2 avr. 2024;46(4):446–70.
 25. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial Intelligence in Undergraduate Medical Education: a scoping review. *Acad Med J Assoc Am Med Coll* 1 nov. 2021;96(115):562–70.
 26. Levin G, Horesh N, Brezinov Y, Meyer R. Performance of ChatGPT in medical examinations: a systematic review and a meta-analysis. *BJOG Int J Obstet Gynaecol*. 2024;131(3):378–80.
 27. Haig A, Dozier M. BEME Guide 3: systematic searching for evidence in medical education - part 2: constructing searches. *Med Teach*. 2003;25:463–84.
 28. Hammick M, Dornan T, Steinert Y. Conducting a best evidence systematic review. Part 1: from idea to data coding. *BEME Guide 13 Med Teach*. 2010;32:3–15.
 29. Kolaski K, Logan LR, Ioannidis JPA. Guidance to best tools and practices for systematic reviews. *Syst Rev* 8 juin. 2023;12(1):96.
 30. Feigerlova E. A Systematic Review on Impact of artificial Intelligence and ChatGPT in training and assessment in health professions education. 4 juin 2024; Disponible sur: Retrieved from <https://osf.io/v5cgp/>.
 31. Collins G, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55–63.
 32. Bates DW, Auerbach A, Schulam P, Wright A, Saria S. Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence. *Ann Intern Med*. 2 juin. 2020;172(11_Supplement):S137–44.
 33. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 24 mars. 2010;340:c332.
 34. Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 1 mars. 2020;2(2):e200029.
 35. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 28 août. 2019;366:l4898.
 36. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
 37. von Elm E, Altman D, Egger M, Pocock P SJ, Gotsche PC, Vandenbroucke JP, et al. The strengthening the reporting of Observational studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147:573–7. Erratum in: *Ann Intern Med*. 2008;148:168.
 38. Schünemann HJ, Brennan S, Akl EA, Hultcrantz M, Alonso-Coello P, Xia J, et al. The development methods of official GRADE articles and requirements for claiming the use of GRADE – a statement by the GRADE guidance group. *J Clin Epidemiol*. 2023;159:79–84.
 39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 29 mars. 2021;372:n71.
 40. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368:l6890.
 41. Gordon M, Gibbs T. STORIES statement: publication standards for health-care education evidence synthesis. *BMC Med* 3 sept. 2014;12(1):143.
 42. Herrington J, Reeves TC, Oliver R. A practical guide to authentic E-learning. New York, NY: Routledge; 2009.
 43. Herrington J. Authentic e-learning in higher education: Design principles for authentic learning environments and tasks. Faculty of Education - Papers. 2006.
 44. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educational Researcher*. 1989;18(1):32–42.
 45. An P, Wang Z. Application value of an artificial intelligence-based diagnosis and recognition system in gastroscopy training for graduate students in gastroenterology: a preliminary study. *Wien Med Wochenschr* 1946. 2023. <https://libezproxy.dundee.ac.uk/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=37676426&site=ehost-live&scope=site>.
 46. Chassagnon G, Billet N, Rutten C, Toussaint T, Cassius de Linval Q, Collin M, et al. Learning from the machine: AI assistance is not an effective learning tool for resident education in chest x-ray interpretation. *Eur Radiol Nov*. 2023;33(11):8241–50.
 47. Furlan R, Gatti M, Menè R, Shiffer D, Marchiori C, Gajj Levra A, et al. A Natural Language Processing-based virtual patient Simulator and Intelligent Tutoring System for the clinical diagnostic process: Simulator Development and Case Study. *JMIR Med Inf* 9 avr. 2021;9(4):e24073.
 48. Kanazawa A, Fujibayashi K, Watanabe Y, Kushi S, Yanagisawa N, Fukataki Y, et al. Evaluation of a medical interview-assistance system using Artificial Intelligence for Resident Physicians interviewing simulated patients: a crossover, Randomized, Controlled Trial. *Int J Environ Res Public Health*. 2023;19(12):6176.
 49. Kim AR, Park AY, Song S, Hong JH, Kim K. A microlearning-based self-directed Learning Chatbot on Medication Administration for New nurses: a feasibility study. *CIN Comput Inf Nurs*. 2024;42(5). https://journals.lww.com/cinjournal/fulltext/2024/05000/a_microlearning_based_self_directed_learning.6.aspx.
 50. Liaw SY, Tan JZ, Bin Rusli KD, Ratan R, Zhou W, Lim S, et al. Artificial Intelligence Versus Human-controlled doctor in virtual reality Simulation for Sepsis Team Training: Randomized Controlled Study. *J Med Internet Res*. 2023b;25:e47748.
 51. Liaw SY, Tan JZ, Lim S, Wentao Zhou, Yap J, Ratan R, et al. Artificial intelligence in virtual reality simulation for interprofessional communication training: mixed method study. *Nurse Educ Today*. 2023a;122:105718.
 52. Meetschen M, Salhöfer L, Beck N, Kroll L, Ziegenfuß CD, Schaarschmidt BM et al. AI-Assisted X-ray Fracture Detection in Residency Training: Evaluation in Pediatric and Adult Trauma Patients. *Diagnostics (Basel)*. 11 mars. 2024;14(6):596(6):596.
 53. Shorey S, Ang ENK, Ng ED, Yap J, Lau LST, Chui CK et al. Evaluation of a theory-based virtual counseling application in nursing education. *CIN Comput Inf Nurs*. 2023;41(6). Disponible sur: https://journals.lww.com/cinjournal/fulltext/2023/06000/evaluation_of_a_theory_based_virtual_counseling.4.aspx.

54. Truong H, Qi D, Ryason A, Sullivan AM, Cudmore J, Alfred S, et al. Does your team know how to respond safely to an operating room fire? Outcomes of a virtual reality, AI-enhanced simulation training. *Surg Endosc Mai*. 2022;36(5):3059–67.
55. Yang Y, Shulruf B. Expert-led and artificial intelligence (AI) system- assisted tutoring course increase confidence of Chinese medical interns on suturing and ligature skills: prospective pilot study. *J Educ Eval Health Prof*. 2019;16(1):7.
56. Zech JR, Ezuma CO, Patel S, Edwards CR, Posner R, Hannon E et al. Artificial intelligence improves resident detection of pediatric and young adult upper extremity fractures. *Skeletal Radiol*. 2024. Disponible sur: <https://doi.org/10.1007/s00256-024-04698-0>.
57. Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2020;122:142–52.
58. Ross JG, Latz E, Meakim CH, Mariani B. TeamSTEPS curricular-wide integration: baccalaureate nursing students' knowledge, attitudes, and perceptions. *Nurse Educ*. 2021;46(6). Disponible sur: https://journals.lww.com/nurseeducatoronline/fulltext/2021/11000/teamsteps_curricular_wide_integration_11.aspx.
59. Bandura A. *Self-efficacy: the Exercise of Control*. New York, NY: Macmillan; 1997.
60. Kiili K. Content creation challenges and flow experience in educational games: the IT-emperor case. *Internet High Educ*. 2005;8(3):183–98.
61. Weng X, Ye H, Dai Y, Ng Olam. Integrating Artificial Intelligence and Computational thinking in Educational contexts: a systematic review of Instructional Design and Student Learning outcomes. *J Educ Comput Res* 1 oct. 2024;62(6):1640–70.
62. Bandura A. *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall; 1977.
63. Torre DM, Daley BJ, Sebastian JL, Elnicki DM. Overview of current learning theories for medical educators. *Am J Med*. 2006;1(10):903–7.
64. Vygotsky L. *Mind in Society: the development of higher psychological processes*. Harvard University Press; 1978.
65. Kolb D. *Experiential Learning: Experience As The Source Of Learning And Development*. 1984.
66. Caffarella RS, Barnett BG. Characteristics of adult learners and foundations of experiential learning. In: Jackson L, Caffarella RS, editors. *Experiential learning: a new approach*. San Francisco, CA: Jossey-Bass Inc; 1994. pp. 29–42.
67. Knowles MS. *Andragogy in action: applying principles of adult learning*. San Francisco: Jossey-Bass; 1984.
68. Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education - a scoping review. *BMC Med Educ*. 2024;24(1):788.
69. Melton R. Objectives. Competencies and Learning Outcomes: Developing Instructional Materials in Open and Distance Learning. 1997. p. 156.
70. Linn RL, Miller MD. *Measurement and Assessment in Teaching*. 9th ed. Upper Saddle River, NJ: Pearson; 2005.
71. Vickers R, Field J, Melakoski C. Media culture 2020: collaborative teaching and blended learning using social media and cloud-based technologies. *Contemp Educational Technol*. 2015;6:62–73.
72. Obilor EI. Feedback and students' learning. *Int J Innov Res Educ* 2 avr. 2019;7:40–7.
73. Jacobs SM, Lundy NN, Issenberg SB, Chandran L. Reimagining Core Entrustable Professional activities for Undergraduate Medical Education in the era of Artificial Intelligence. *JMIR Med Educ*. 2023;9:e50903.
74. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Sci* 25 oct. 2019;366(6464):447–53.
75. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Min Knowl Discov* 1 Juill. 2019;9(4):e1312.
76. Markovski Y. Data usage for consumer services FAQ. OpenAI. juill 2024; Disponible sur: <https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq>.
77. Reilly C, Reeves TC. Refining active learning design principles through design-based research. *Act Learn High Educ* 1 mars. 2024;25(1):81–100.
78. Harden RM. Learning outcomes and instructional objectives: is there a difference? *Med Teach*. 2002;24(2):151–5.
79. Hartel R, Foegeding EA. Learning: objectives, competencies, or outcomes ? *J Food Sci Educ*. 2004;3:69–70.
80. Bandura A. *Social foundations of thought and action: A social cognitive theory*. 1986. p. 640.
81. Lin CC, Huang A, Lu O. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learn Environ*. 2023;10(41).
82. Chan KS, Zary N. Applications and Challenges of Implementing Artificial Intelligence in Medical Education: integrative review. *JMIR Med Educ* 15 juin. 2019;5(1):e13930.
83. Masters K. Artificial intelligence in medical education. *Med Teach*. 2019;41(9):976–80.
84. Masters K, Herrmann-Werner A, Festl-Wietek T, Taylor D. Preparing for Artificial General Intelligence (AGI) in Health professions Education: AMEE Guide 172. *Med Teach*. 2024;1–14.
85. Daneshvar N, Pandita D, Erickson S, Sulmasy LS, DeCamp M. Artificial Intelligence in the provision of Health Care: an American College of Physicians policy position paper. *Ann Intern Med*. 4 juin 2024 [cité 22 juin 2024]; Disponible sur: <https://doi.org/10.7326/M24-0146>.
86. Charow R, Jeyakumar T, Younus S, Dolatabadi E, Sahlia M, Al-Mouaswas D, et al. Artificial intelligence education programs for health care professionals: scoping review. *JMIR Med Educ*. 2021;7(4):e31043.
87. Ng FYC, Thirunavukarasu AJ, Cheng H, Tan TF, Gutierrez L, Lan Y et al. Artificial intelligence education: an evidence-based medicine approach for consumers, translators, and developers. *Cell Rep Med [Internet]* 17 oct 2023 [cité 27 déc 2024];4(10). Disponible sur: <https://doi.org/10.1016/j.xcrm.2023.101230>.
88. Száva-Kováts E. Unfounded attribution of the half-life index-number of literature obsolescence to Burton and Kebler: A Literature Science Study. *J Am Soc Inf Sci Technol*. 2002;53:1098–105.
89. Wartman S, Combs C. Medical Education must move from the information age to the age of Artificial Intelligence. *Acad Med* 1 nov. 2017;93:1.
90. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn Environ*. 22 févr. 2023;10(1):15.
91. McKenney S, Reeves TC. Educational design research: portraying, conducting, and enhancing productive scholarship. *Med Educ* 1 janv. 2021;55(1):82–92.
92. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well- built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12e3.
93. Frandsen TF, Bruun Nielsen MF, Lindhardt CL, Eriksen MB. Using the full PICO model as a search tool for systematic reviews resulted in lower recall for some PICO elements. *J Clin Epidemiol* 1 nov. 2020;127:69–75.
94. Flanagan A, Pirracchio R, Khera R, Berkwitz M, Hswen Y, Bibbins-Domingo K. Reporting Use of AI in Research and Scholarly Publication—JAMA Network Guidance. *JAMA* 2 avr. 2024;331(13):1096–8.

Publisher's note

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