Revista Española
de Educación Médica

UNIVERSIDAD DE MURCIA

Mapping Applications and Outcomes of Large-Language-Model-Generated Cases in Health Professions Education: A Scoping Review.

Abdullah Bedir Kaya[1,2], Emre Emekli[3], Yavuz Selim Kıyak[4]*

[1]Department of Computer Technologies, Hitit University, Çorum 19030, Turkiye, abedirkaya@hitit.edu.tr, https://orcid.org/0000-0003-0124-2127
[2]Health Sciences Institute, Gazi University, Ankara 06500, Turkiye.
[3]Department of Radiology, Faculty of Medicine, Eskişehir Osmangazi University, Eskişehir, Turkiye, eemekli@ogu.edu.tr, https://orcid.org/0000-0001-5989-1897
[4]Department of Medical Education and Informatics, Faculty of Medicine, Gazi University, Ankara 06500, Turkiye, yskiyak@gazi.edu.tr, https://orcid.org/0000-0002-5026-3234
[*] Correspondence: yskiyak@gazi.edu.tr

Submitted: 12/5/25; Accepted: 12/29/25; Published: 1/2/26

Abstract.

Objective: Large language models (LLMs) have rapidly permeated health professions education and are increasingly used to generate clinical cases and vignettes, yet their characteristics, evaluation methods, and educational impact remain unclear. To map how LLMs are used to generate cases in health professions education and to summarize reported case characteristics, evaluation approaches, bias, and educational outcomes. **Methods:** We conducted a scoping review following Arksey and O'Malley's framework and reported using PRISMA-ScR. PubMed, Web of Science, and Scopus were searched on 27 August 2025. Of 2023 records, 72 full texts were assessed and 23 studies met inclusion criteria. Data were charted with a structured extraction form. **Results:** Across the 23 studies, 33 distinct LLMs were used, most commonly GPT-based models (54.5%). Cases were mainly text-based (69%), with additional image- (20.7%) and audio-based (10.3%) formats across 23 clinical and educational domains. Prompts were reported in 65.2% of studies, and 60.9% included a formal quality evaluation, ranging from high quality to clearly problematic examples. Seven studies (30.4%) identified bias or discriminatory patterns. Student participation occurred in 39.1% of studies, but no higher-level educational outcomes such as behavior change or long-term performance were reported. **Conclusions:** LLM-generated cases appear feasible and versatile across health professions education but are supported by early, methodologically heterogeneous evidence. Future research should standardize quality evaluation, rigorously assess learning and behavioral outcomes, and systematically audit bias in generated content.

Keywords: artificial intelligence, large language models, case generate, medical education

1. Introduction

Generative artificial intelligence (AI) has rapidly permeated medical education; ChatGPT, having reached 100 million users within two months of its release and becoming the fastest-growing digital application in history (1), has simultaneously driven a remarkable surge in scholarly output, with monthly publications on this topic increasing from just 2 in March 2023 to 33 by May 2025 (2). AI is being used in health professions education by supporting preclinical learning, enabling innovative teaching methodologies, simulating clinical environments, and contributing to assessment processes for medical examinations (2). Within this wider landscape, one of the most intuitive uses of LLMs is the generation of clinical cases and vignettes, core pedagogical tools in health professions education.

Clinical cases and vignettes have long been central to health professions education. Case-based learning is an effective instructional method that enhances academic performance and case analysis skills in health professions education (3-4). Because cases are usually constructed manually by content experts, they demand considerable time and pedagogical expertise, and their quality depends on how well they integrate basic science, clinical reasoning, and contextual factors. At this point, LLMs can be used for case generation. This scoping review aims to systematically map how LLMs are being applied to generate clinical cases in health professions education and to reveal the associated educational outcomes.

2. Methods

We conducted this scoping review following the methodological framework proposed by Arksey and O'Malley (5) for scoping studies. While systematic reviews remain the predominant method for synthesizing evidence in medical education research (6), the exploratory and wide-ranging scope of our question, without a clearly defined hypothesis, made a scoping review design more appropriate for our study. The review process was reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist to ensure methodological transparency and reproducibility.

Our study was guided by a post-positivist paradigm, acknowledging that while an objective synthesis of published data is sought, complete objectivity is unattainable due to the interpretive nature of data extraction and synthesis. This epistemological stance aligns with contemporary perspectives in health professions education research (7). Following the framework (5), the review was conducted in five sequential stages:

- Identifying the research question
- Identifying relevant studies
- Study selection
- Charting the data
- Collating, summarizing, and reporting the results

Identifying the Research Question

We adopted the definition of a "case" as a structured scenario-rooted in patient cases that may be real, simulated, virtual, or text-based designed to facilitate learning, assessment, or reflective practice in health professions education (4). For the purposes of this review, a Large Language Model (LLM)-generated case was operationally defined as any educational case, scenario, or vignette, of any type including text, image, or audio, produced wholly by a large language model and intended for using only as a case, not intended for other combined purposes such as case-based multiple-choice questions and interactive virtual patients, within a health professions education context. Based on this definition, our primary research question was formulated as follows: How are LLM's being used to generate cases in health professions education, and what characteristics and educational outcomes are reported in these studies?

Identifying Relevant Studies

After several iterative refinements, a comprehensive search was conducted in PubMed, Web of Science and Scopus on 27 August, 2025, using the following query, with no date restrictions applied: ("medical" AND ("education" OR "student" OR "training" OR "curriculum") OR ("health professions education" OR "nursing education" OR "dental education" OR "pharmacy education" OR "veterinary education")) AND ("large language model" OR "artificial intelligence" OR "large language models" OR "chatgpt" OR "generative AI") AND ("case" OR "vignette").

Study Selection

All retrieved records were imported into Rayyan AI, an online tool designed to facilitate screening in literature reviews. Duplicate entries were automatically removed. The remaining studies were first screened by one reviewer who examined titles and abstracts according to the inclusion and exclusion criteria. The selected articles were then independently reviewed by two researchers (interrater agreement rate = 88,6%), and in cases of disagreement, a third reviewer acted as an arbiter. The inclusion criteria were as follows: studies that reported the use of LLMs to generate cases (in any modality; text, image, or audio) within the context of health professions education and that generated at least one case. Studies were included if the generated cases were intended for learning, teaching, or assessment purposes. And the exclusion criteria: studies that did not generate any cases; studies explaining only the methodology of case generation without producing cases; those using virtual patients; studies comparing LLM model performance; research focusing on AI-driven clinical decision support systems; studies focusing on case-based multiple-choice question (MCQ) generation. Non-English publications, review papers, preprints, and withdrawn articles were also excluded. Borderline studies were included only when the LLM output constituted a patient-centered scenario/vignette functioning as a 'case'; studies focused primarily on standalone assessment item generation (e.g., MCQs) or virtual patient systems were excluded. Database searches retrieved 2023 records. After removing duplicates, 1306 titles and abstracts were screened by one reviewer, leading to 72 full-text articles assessed for eligibility by at least two reviewers. Conflicts were solved by a third reviewers. Following exclusions, 23 studies met inclusion criteria and were included in the review. The list of excluded 49 reports with reasons was provided as a supplementary material. The selection process is presented in the flow diagram (figure 1).

Charting the Data

Data extraction was conducted using a structured charting form that was iteratively refined after pilot testing on a subset of included studies. The form was designed to systematically capture the context, methodology, and educational implications of each study on LLM-generated case production in health professions education. Two reviewers were charted the data independently. Interrater agreement rate was 88,6%. The third reviewer acted as an arbiter. For each included study, the charting process captured detailed information including the author(s), country, publication type, discipline or department, case type and number, LLM version, case area, information on prompting methodology, and student information. In addition, the

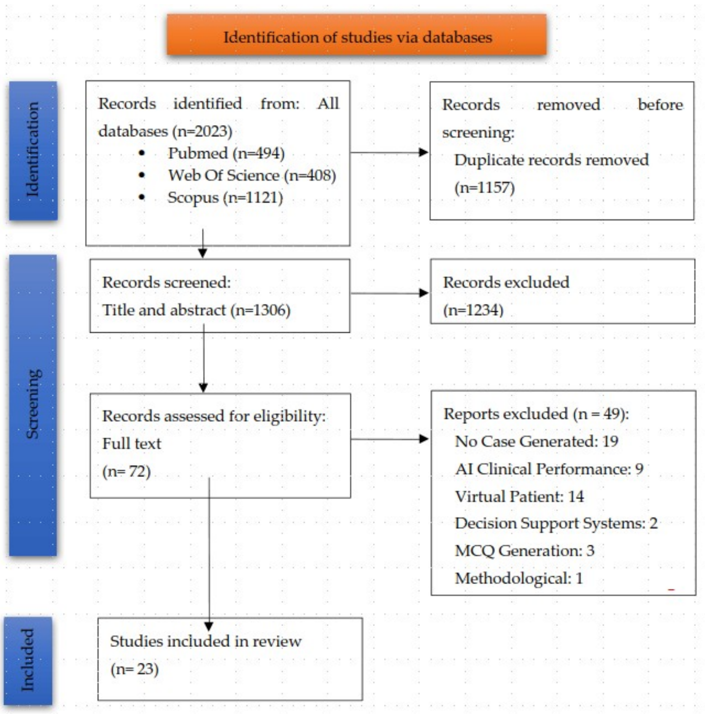

Figure 1. Identification of studies via databases process.

evaluation of the quality of the generated cases and the examination of potential discrimination issues embedded within the cases were also conducted. All extracted data were summarized in tabular form (see Supplementary Table 1) and analyzed descriptively and narratively to map the range of applications, methodological diversity, and outcomes of LLM-generated case studies across health professions education.

Collating, Summarizing, and Reporting the Result

Data were synthesized descriptively and narratively to map the landscape of LLM-generated cases in health professions education. Extracted data were summarized in tables and categorized by geographic distribution, LLM version, study design, case type, and evaluation outcomes. Reported findings were analyzed to identify trends in case quality, educational use, and ethical considerations such as bias and fairness.

3. Results

In total, 23 studies were included (8–30).

Geographic Distribution

Geographically, the United States accounted for the largest proportion of studies (n = 5), followed by Turkiye (n = 3) and China (n = 3). Other contributing countries included Japan (n = 2), Iran, India, Grenada, Saudi Arabia, Netherlands, Mexico, Spain, Canada, Bahrain, and the United Kingdom, each represented by one study. When categorized by global regions, 10 studies (43.5%) originated from the Global North, while 13 studies (56.5%) came from the Global South.

LLM Versions, Case Types, and Prompting Strategies

Across the 23 included studies, a total of 33 distinct LLMs were employed for case generation. Of these, 18 (54.5%) belonged to the OpenAI family (e.g., GPT-3.5, GPT-4), reflecting its dominant role. Google Gemini/Bard was utilized in 3 studies (9.1%), DALL·E in 2 studies (6.1%), while 10 other models (30.3%), including various domain-specific and open-source systems, were each used only once. Across the included studies, 20 text-based (69.0%), 6 image-based (20.7%), and 3 audio-based (10.3%). In 5(16,5%) studies, more than one modality of case generation was used. While text-based dominated, some studies combined text, image, and audio in multimodal formats. Also number of generated cases per study ranged from 1 to 18,000 (29), with a median of approximately 30 cases. A total of 15 studies (65.2%) disclosed their prompts, 11 (47.8%) in the main text and 4 (17.4%) in supplementary materials, while 8 (34.8%) did not. Iterative prompting was used in 9 studies (39.1%), and 14 (60.9%) employed a single-prompt approach.

Study Context, Discipline, and Student Involvement

All studies were original research articles, primarily conducted in medicine 17 (73.9%), followed by nursing 5 (21.7%) and veterinary medicine 1 (4.3%), highlighting a strong emphasis on medical education. LLM-generated cases covered a broad range of 23 areas within health professions education. The most frequent domains were clinical health (n = 10, 43.5%), including cardiology(16), endocrinology(10,20), oncology(24,30), infectious diseases(27), and surgical care, and nursing-related contexts (n = 6, 26.1%), such as medical-surgical nursing(18,19,25), pharmacology(26), and nursing care(8,17). Additionally, communication (11) and professional skills(22) were explored in 2 studies (8.6%), while genetic and metabolic disorders(9) (e.g., β-thalassemia, cystic fibrosis, Tay–Sachs disease, aldehyde dehydrogenase deficiency) appeared in 1 studies (4.3%). Overall, the cases represented a diverse mix of clinical, educational, and ethical topics, highlighting LLMs' adaptability across multiple disciplines and instructional settings. Student participation occurred in 9 studies (39.1%), involving cohorts ranging from 9 to 251 students across medical, nursing, radiography, and veterinary programs, mostly at advanced undergraduate levels.

Case Quality and Bias Evaluation

Formal evaluation of LLM-generated case quality was reported in 14 (60.9%), whereas 9 (39.1%) did not include a formal assessment. Among the evaluated studies, 8 (57.1%) used Likert-scale questionnaires or structured forms, 3 (21.4%) employed statistical or model-based analyses, and 3 (21.4%) relied on expert panel or faculty reviews. Several studies rated the cases as high quality, noting strong alignment with learning objectives, completeness, or diagnostic coherence. Expert evaluations ranged broadly, with cases described across categories from high quality (20, 28) to problematic (9, 16).

Also 7 (30,4%) studies (9, 13, 20–23, 29) reported the presence of bias in LLM-generated cases. The identified issues included reinforcement of ethnic or ethnoracial stereotypes (9), associations between demographic characteristics (such as gender, education level, income, insurance status, or nationality) and behavioral assumptions (21) and the presence of low-level discriminatory patterns (23). Any issue explicitly reported or labeled as 'bias' by the included studies was charted and treated as bias in this review.

4. Discussion

This scoping review mapped 23 studies that used large language models to generate cases for health professions education. In contrast to the pronounced Global North–South divide, our findings did not reveal a substantial gap; research activity was distributed across both regions, with only a marginal predominance of studies originating from the Global North. As (31) argues "Despite the growing diversity of the South, these shifts have not overturned the two-tier structure of the global order," a perspective that remains relevant when interpreting our geographic distribution. More than half of all LLMs used were from the ChatGPT family (54.5%), reflecting a strong dependence on OpenAI models similar to patterns documented in other AI-driven educational innovations. A wide range of 33 distinct LLMs were identified, yet the reliance on a few dominant platforms suggests unequal access to emerging technologies, echoing the structural asymmetries highlighted.

Across the included studies, LLMs were used to generate diverse types of cases. Text-based cases remained the most prevalent (69%), although some studies demonstrated early multimodal experimentation by combining text, images, and audio, a diversity that indicates alignment with multimedia learning principles (32). The number of generated cases varied substantially between studies, ranging from a single case to large-scale outputs exceeding 18,000 cases. Numbers demonstrates the substantial potential of LLM-generated cases in health professions training.

The majority of studies (73.9%) were conducted in medical education. The cases spanned 23 different content areas, including clinical medicine, nursing care, communication and professional skills, and rare genetic and metabolic disorders. The breadth of subject matter suggests that LLM-generated cases can be used across nearly all domains of health professions education. Also most studies reported their case-generation prompts, an important practice for ensuring reproducibility.

Most generated cases received formal evaluation, with quality ratings ranging from excellent to problematic. These findings highlight the need for establishing a clear, standardized formal evaluation process for LLM-generated cases in health professions education. Seven studies reported bias-related issues in LLM-generated cases. These included ethnic or ethnoracial stereotypes and links made between demographic characteristics and assumed behaviors. These findings suggest that bias assessment and mitigation should be considered a core quality domain of LLM-generated cases, not an optional add-on, prior to deployment with learners. Such findings are consistent with broader concerns that LLMs may reproduce or amplify existing biases in their training data. Additionally, one study reported that 70% of Japanese-generated cases were rated as acceptable, whereas 27% required revision, attributing this largely to the fact that many LLMs are predominantly trained on English-language corpora (28). Therefore, greater caution is warranted when generating cases in languages other than English, particularly to address linguistic fidelity and cultural appropriateness.

Despite the diversity of applications and reported outcomes, the educational impact of LLM-generated cases could not be examined (33). This situation reflects the early developmental stage of LLM-generated cases.

Future studies should address several critical areas: (a) evaluating higher-level educational outcomes, (b) examining whether LLM-generated cases contain discriminatory patterns, and (c) developing LLM-generated cases that align with multimedia learning principles through text-image-audio-video combinations. Additionally, (d) transparent reporting of the prompting process and model parameters will be essential to ensure reproducibility and methodological rigor.

This review has several limitations. First, the search strategy was limited to English-language, peer-reviewed sources, thereby excluding grey literature and preprints. This language restriction may disproportionately overlook innovative work in regions where alternative LLMs (e.g., DeepSeek, Qwen, Kimi, GLM) are rapidly evolving. Additionally, excluding grey literature and preprints may have introduced publication and timeliness bias. Second, as a scoping review, no risk-of-bias appraisal was conducted, leaving the methodological quality of the included studies uncertain. Therefore, our findings should be interpreted as descriptive mapping of the literature; the absence of a formal risk-of-bias assessment limits the validity of effectiveness-related conclusions. The search was conducted on 27 August 2025, which means that recently published studies. Particularly those involving more advanced LLMs such as ChatGPT 5 and Gemini 3 Pro, may not have been captured. Additionally, the scope was limited to studies that directly generated cases, excluding case-based multiple-choice question generation and virtual patient interactions; therefore, related but distinct LLM-based instructional approaches fall outside this synthesis. Additionally, the initial title/abstract screening was conducted by a single reviewer, which may have introduced selection bias despite subsequent independent full-text review.

Future studies should rigorously examine the educational impact by exploring how LLM-generated cases influence learning processes, skill development, and real-world clinical performance. In addition, evaluations of bias and discrimination are essential to ensure safe and ethical use. Leveraging multimodal case-generation, such as integrating text, image, audio, or video, may help align these tools with contemporary learning principles and deepen pedagogical value.

5. Conclusions

- LLM-generated cases are increasingly being used in health professions education, but the field is still in an early developmental stage. The 23 mapped studies show both high-quality and problematic examples. Therefore, transparency in prompting and methodological clarity are essential.
- Studies reported overall feasibility and generally positive evaluations of case quality, yet concerns persist regarding structural inaccuracies, inconsistent quality-assessment methods, and the presence of bias or discriminatory patterns in some generated cases.
- Among the mapped studies, no higher-level educational outcomes of the use of LLM-generated cases, such as behavior change or long-term performance, were reported,

indicating that warrants further investigation.

Annex 1: Supplementary Table 1 (included studies extract data).

Funding: No funding received for this study.

Declaration of conflict of interest: The authors declare that they have no conflict of interest.

Data availability: The datasets generated and analyzed during the current study are available in the Zenodo repository at 10.5281/zenodo.17830543

Author contributions: ABK, EE, YSK; Data collection: ABK, EE; Formal analysis: ABK, EE, YSK; Methodology: ABK, EE, YSK; Roles/Writing: - original draft: ABK, YSK; and Writing - review & editing: ABK, EE, YSK.

6. References

1. Krystal Hu. ChatGPT sets record for fastest-growing user base - analyst note. Reuters **2023**. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/
2. Zhang Y, Xie X, Xu Q. ChatGPT in medical education: Bibliometric and visual analysis. *JMIR Med Educ.* **2025**, 11, e72356–e72356. https://doi.org/10.2196/72356.
3. Tsekhmister Y. Effectiveness of case-based learning in medical and pharmacy education: A meta-analysis. Electron *J Gen Med.* **2023**, 20(5), em515. https://doi.org/10.29333/ejgm/13315.
4. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical Teacher* **2012**, 34, e421–e444. https://doi.org/10.3109/0142159X.2012.680939.
5. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International *Journal of Social Research Methodology* **2005**, 8, 19–32. https://doi.org/10.1080/1364557032000119616.
6. Maggio LA, Samuel A, Stellrecht E. Systematic reviews in medical education. *Journal of Graduate Medical Education* **2022**, 14, 171–175. https://doi.org/10.4300/JGME-D-22-00113.1.
7. Thomas A, Lubarsky S, Varpio L, Durning SJ, Young ME. Scoping reviews in health professions education: challenges, considerations and lessons learned about epistemology and methodology. *Adv in Health Sci Educ.* **2020**, 25, 989–1002. https://doi.org/10.1007/s10459-019-09932-2.
8. Akutay S, Yüceler Kaçmaz H, Kahraman H. The effect of artificial intelligence supported case analysis on nursing students' case management performance and satisfaction: A randomized controlled trial. *Nurse Education in Practice* **2024**, 80, 104142. https://doi.org/10.1016/j.nepr.2024.104142.
9. Andrew K, Montalbano MJ. Through a Glass Darkly: Perceptions of Ethnoracial Identity in Artificial Intelligence Generated Medical Vignettes and Images. *Medical Science Educator* **2025**, 35, 1473–1488. https://doi.org/10.1007/s40670-025-02332-9.
10. Arain SA, Akhund SA, Barakzai MA, Meo SA. Transforming medical education: Leveraging large language models to enhance PBL-a proof-of-concept study. *Advances in Physiology Education* **2025**, 49, 398–404. https://doi.org/10.1152/advan.00209.2024.
11. Artemiou E, Hooper S, Dascanio L, Schmidt M, Gilbert G. Introducing AI-generated cases (AI-cases) & standardized clients (AI-SCs) in communication training for veterinary students: Perceptions and adoption challenges. *Front Vet Sci.* **2025**, 11, 1504598. https://doi.org/10.3389/fvets.2024.1504598.
12. Aygün İ, Kaya M. Use of large language models for medical synthetic data generation in mental illness. *IET Conf Proc.* **2024**, 652–656. https://doi.org/10.1049/icp.2024.1033.
13. Bakkum MJ, Hartjes MG, Piet JD, Donker EM, Likic R, Sanz E, et al. Using artificial intelligence to create diverse and inclusive medical case vignettes for education. *British Journal Of Clinical Pharmacology* **2024**, 90, 640–648. https://doi.org/10.1111/bcp.15977.

14. Benoit JRA. ChatGPT for clinical vignette generation, revision, and evaluation. *MedRxiv* **2023.** https://doi.org/10.1101/2023.02.04.23285478.
15. Coşkun Ö, Kıyak YS, Budakoğlu Iİ. ChatGPT to generate clinical vignettes for teaching and multiple-choice questions for assessment: A randomized controlled experiment. *Medical Teacher* **2025**, 47, 268–274. https://doi.org/10.1080/0142159X.2024.2327477.
16. Ghaffari F, Langarizadeh M, Nabovati E, Sabery M. Effectiveness of ChatGPT for clinical scenario generation: A qualitative study. *Archives Of Academic Emergency Medicine* **2025**, 13, e49. https://doi.org/10.22037/aaemj.v13i1.2690.
17. Higashitsuji A, Otsuka T, Watanabe K. Impact of ChatGPT on case creation efficiency and learning quality in case-based learning for undergraduate nursing students. *Teaching and Learning in Nursing* **2025**, 20, e159–166. https://doi.org/10.1016/j.teln.2024.10.002.
18. Mondal H, Marndi G, Behera J, Mondal S. ChatGPT for teachers: Practical examples for utilizing artificial intelligence for educational purposes. *Indian Journal of Vascular and Endovascular Surgery* **2023**, 10, 200–205. https://doi.org/10.4103/ijves.ijves_37_23.
19. Jackson ML. From cases to confidence: Developing diagnostic reasoning skills through collaborative learning in graduate nursing education. *Nurs Educ Perspect.* **2025**, 46, 319–321. https://doi.org/10.1097/01.NEP.0000000000001438.
20. Lam G, Shammoon Y, Coulson A, Lalloo F, Maini A, Amin A, et al. Utility of large language models for creating clinical assessment items. *Medical Teacher* **2025**, 47, 878–882. https://doi.org/10.1080/0142159X.2024.2382860.
21. Liu C, Zheng J, Liu Y, Wang X, Zhang Y, Fu Q, et al. Potential to perpetuate social biases in health care by Chinese large language models: A model evaluation study. *Int J Equity Health* **2025**, 24, 206. https://doi.org/10.1186/s12939-025-02581-5
22. Lopez M, Goh P-S. Catering for the Needs of Diverse Patient Populations: Using ChatGPT to Design Case-Based Learning Scenarios. *Med Sci Educ.* **2024**, 34, 319–325. https://doi.org/10.1007/s40670-024-01975-4.
23. Rao AS, Kim J, Mu A, Young CC, Kalmowitz E, Senter-Zapata M, et al. Synthetic medical education in dermatology leveraging generative artificial intelligence. *Npj Digit Med.* **2025**, 8, 247. https://doi.org/10.1038/s41746-025-01650-x.
24. Ruiz Sarrias O, Martínez Del Prado MP, Sala Gonzalez MÁ, Azcuna Sagarduy J, Casado Cuesta P, Figaredo Berjano C, et al. Leveraging large language models for precision monitoring of chemotherapy-induced toxicities: A pilot study with expert comparisons and future directions. *Cancers* **2024**, 16, 2830. https://doi.org/10.3390/cancers16162830.
25. Silvestri-Elmore A, Burton C. How can nursing faculty create case studies using AI and educational technology? *Nurse Educator* **2025**, 50, 35–39. https://doi.org/10.1097/NNE.0000000000001734.
26. Sridharan K, Sequeira RP. Evaluation of artificial intelligence-generated drug therapy communication skill competencies in medical education. *Br J Clin Pharmacol.* **2025**, 91, 2168–2175. https://doi.org/10.1111/bcp.16144.
27. Xie W, Yuan Z, Si Y, Huang Z, Li Y, Wu F, et al. Enhancing medical students' diagnostic accuracy of infectious keratitis with AI-generated images. *BMC Medical Education* **2025**, 25, 1027. https://doi.org/10.1186/s12909-025-07592-y.
28. Yanagita Y, Yokokawa D, Uchida S, Li Y, Uehara T, Ikusaka M. Can AI-Generated clinical vignettes in japanese be used medically and linguistically? *J Gen Intern Med.* **2024**, 39, 3282–3289. https://doi.org/10.1007/s11606-024-09031-y.
29. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. *The Lancet Digital Health* **2024**, 6, e12–22. https://doi.org/10.1016/S2589-7500(23)00225-X.
30. Zhong D, Chow SKK. Investigating the potential of generative AI clinical case-based simulations on radiography education: A pilot study. *Journal of Imaging Informatics in Medicine* **2025**, 1–13. https://doi.org/10.1007/s10278-025-01601-8.

31. Lees N. The Brandt Line after forty years: The more North–South relations change, the more they stay the same? *Rev Int Stud.* **2021**, 47(1), 85–106. https://doi.org/10.1017/S026021052000039X.
32. Mayer R. Multimedia Learning. 3rd ed. Cambridge University Press; **2020**. https://doi.org/10.1017/9781316941355.
33. Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Med Educ.* **2025**, 25, 129. https://doi.org/10.1186/s12909-025-06719-5.