Author(s)	Country	Publication type	Department	Case Type	Case Number	LLM version	Is it aRCT?	Area and/orTopic	Is prompt given?	Iterative?	Student Participation	Students Information	Formal evaluation of Case Quality based on	Case Quality (formal assesment outcome)	Bias
Jackson,	USA	Orginal Article	Nursing	Text	260	Chat GPT	No	Nursing	No	No	Yes	17 first-year advanced practice provider students	No	No	No
Silvestri - Elmore et al,	USA	Orginal Article	Nursing	Text	unspecified	Chat GPT	No	Medical - surgical nursing, Pharmacology	No	Yes	Yes	Prelicensure nursing students (n=not specified)	No	No	No
Lopez et al,	Mexico	Orginal Article	Medicine	Text	5	Chat GPT	No	Cultural-Sensetive MEDU	Yes	Yes	No		No	No	Yes (Cases were generated based on religion, language, gender, culture, and socioeconomic inequalities.)
Himel et al,	India	Orginal Article	Medicine	Text	1	Chat GPT	No	Endovascular surgery	No	No	No		No	No	No
Benoit	Canada	Orginal Article	Nursing	Text	45	Chat GPT	No	Childhood illness	Yes	Yes	No		No	No	No
Aygun et al.	Turkiye	Conference	Medicine	Text	200	Chat GPT - Google Bard	No	Mental Illness	No	No	No		Yes (a quantitative assessment by a pre-trained BERT model, which measured its accuracy in detecting psychological illness within the synthetic medical texts.) - MODEL BASED	The quality of the AI-generated cases was high, as a pre-trained BERT model identified psychological illness in the synthetic texts with accuracy rates (89% and 87%) that were comparable to on real patient data.	No
Higashitsuji et al,	Japan	Orginal Article	Nursing	Text	6	Chat GPT 3.5	No	Nursing	No	No	Yes	9 fourth-year nursing students	Yes (Assessed using direct time measurement for efficiency, a 5-point Likert scale questionnaire for faculty burden, and the Method for Visualizing Group Work Activities for objective and subjective discussion quality) - LIKERT	No significant differences in the quality of group discussions, indicating that using ChatGPT for case creation did not negatively affect the learning quality and using ChatGPT directly improves the quality of education by increasing faculty members' time in contemplating cases	No
Andrew et al.	Grenada	Orginal article	Medicine	Text - Image	200 text based case, does not specify the total number of visual cases	Chat GPT 3.5 - Bing Image Creator	No	Sickle cell anemia Beta-thalassemia Cystic fibrosis Tay-Sachs disease Aldehyde dehydrogenase	Yes	Yes	No		Yes (based on comparing the ethnoracial representation in the generated texts against epidemiological data, and the representation in the images against human rater assessments.) - EXPERT PANEL	Case quality is problematic as it reinforces ethnic and racial stereotypes and biases.	Yes (AI strongly reinforces and exaggerates ethnoracial stereotypes in both its text-based vignettes and visual outputs.)
Bakkum et al.	Netherlands	Orginal Article	Medicine	Text - Image	30 text - 4 Image	Chat GPT 3.5, Adobe Firefly beta	No	-	Yes	Yes	No		No	No	Yes (While AI offers the potential to mitigate stereotypes by generating diverse cases , it concurrently reproduces inherent biases from its training data, such as favoring young and white individuals or creating new stereotypical associations (e.g., linking the occupation 'cook' with a high BMI and French ethnicity with 'wine connoisseur')

Author(s)	Country	Publication type	Department	Case Type	Case Number	LLM version	Is it aRCT?	Area and/orTopic	Is prompt given?	Iterative?	Student Participation	Students Information	Formal evaluation of Case Quality based on	Case Quality (formal assesment outcome)	Bias
Artemiou et al.	USA	Orginal Article	Veterinary	Text - Image - Sound	6	Chat GPT 3.5, Talkr Live	No	Communucation Skills / Clinical interview	Yes	No	Yes	234 veterinery students (second semester=102, fourth semester=71, sixth semester=61)	No	No	No
Zack et al,	USA	Orginal Article	Medicine	Text	18.000	Chat GPT 4	No	18 different disease	Suplementary materials	No	No		Yes (This model evaluation study formally assessed GPT-4 for racial and gender biases across four clinical applications by using standard statistical tests to analyze its generated outputs and responses to clinical vignettes.) - MODEL BASED	The evaluation found that GPT-4-generated clinical vignettes failed to model true demographic diversity and exaggerated stereotypes , while its assessment of standardized cases perpetuated significant racial and gender biases in diagnostic reasoning and treatment recommendations.	Yes
Lam et al,	UK	Orginal Article	Medicine	Text	2	Chat GPT 4	Yes	Primary Care History and Diabetes/Endocrinology	Suplementary material	Yes	Yes	251 final year medical students	Yes (The cases were also formally assessed by 15 blinded examiners using a five-point Likert scale questionnaire rating their 'realism' and 'difficulty') - LIKERT	100% of the examiners agreed the GPT-assisted cases were realistic in content and appropriately difficult and GPT-assistance resulted in significant labour cost savings, and required substantially less time.	Yes (The study acknowledges the recognized limitations of potential biases in GPT-4 and describes how this was addressed by including equality, diversity, and inclusion (EDI) principles in the prompts and through review by EDI-trained faculty).
Ruiz Sarrias et al,	Spania	Orginal Article	Medicine	Text-Sound	30	Chat GPT 4	No	Oncology	Suplementary materials	Yes	No		No	No	No
Yanagita et al,	Japan	Orginal Article	Medicine	Text	202	Chat GPT 4	No	Japanese medical cases	Yes	No	No		Yes (Three physicians who assessed the AI-generated clinical vignettes for both "medical accuracy" and "Japanese-language accuracy," assigning a score on a five-point scale for each criterion) - LIKERT	The use of a generative AI, GPT, in conjunction with corrections by Japanese physicians enabled the creation of Japanese-language clinical vignettes with 97% medical and linguistic accuracy.	No
Ghaffari et al.	Iran	Orginal Article	Medicine	Text	3	Chat GPT 4	No	Cardiovascular (cardiogenic shock, postoperative cardiac tamponade after heart surgery, heart failure)	Yes	No	No		Yes (A qualitative study using thematic analysis of feedback provided by a five-member expert panel) EXPERT PANEL	While experts found the AI-generated cases to be realistic and aligned with clinical guidelines, they also identified significant errors, inconsistencies, and missing information, necessitating expert review and refinement.	No
Rao et al,	USA	Orginal Article	Medicine	Text	20	Chat GPT 4o	No	skin and soft tissue diseases	Yes	No	No		Yes(Three practicing attending physicians commonly encountering the topics at hand at our institution rated the answers based on how well they fit the criteria using a Likert scale - LIKERT	This work demonstrates that off-the-shelf LLMs like GPT-4 hold great potential to be used for clinical vignette generation for standardized examination and teaching purposes	Yes (A low level of discrimination was observed in the generated cases.)
Sridharan and Sequeira	Bahrain	Orginal Article	Medicine	Text	15	Chat GPT, Google Bard, Poe Assistant	No	Drug threrapy communication skills	Yes	No	No		Yes (Two authors evaluated the AI-generated communication components using a consensus-based approach with a checklist , and assessed the veracity of the instructions and warnings by referencing the British National Formulary (BNF) and Medline Plus.) - EXPERT PANEL	While the cases generated by Poe Assistant and ChatGPT showed broad convergence on salutation, drug name, treatment goals, and follow-up schedules, they exhibited significant differences and deficiencies in critical components like instruction clarity, side effects, warnings, and patient empowerment.	No

Author(s)	Country	Publication type	Department	Case Type	Case Number	LLM version	Is it aRCT?	Area and/orTopic	Is prompt given?	Iterative?	Student Participation	Students Information	Formal evaluation of Case Quality based on	Case Quality (formal assesment outcome)	Bias
Coşkun et al.	Turkiye	Original Article	Medicine	Text	37	ChatGPT-3.5	Yes	Evidence-based medicine training	Yes	Yes	Yes	74 fourth-year medical students (37 AI group, 37 human written group)	Yes (medical students evaluated the cases by rating 10 statements on a 5-point Likert scale to compare the ChatGPT-generated vignettes against the human-written versions.) - LIKERT	ChatGPT-generated vignettes is comparable to those created by human authors, as medical students' evaluations showed no significant difference between the two groups.	No
Zhong et al,	China	Orginal Article	Medical imaging technologists/ radiographers	Image	unspecified	DALL E 3	No	Imaging of brain cancer and heart disease	No	No	Yes	122 radiography students (42 first-year, 39 second-year, and 41 third-year students)	Yes (using pre- and post-lecture surveys, assessed the perceptions of 122 students and 152 graduates regarding an in-house AI-based imaging simulation tool, whose generated images were also evaluated by lecturers.) - LIKERT	These insights suggest that consensus on the educational value of generative AI could substantially enhance radiographer training, particularly in resource-limited regions.	No
Akutay et al.	Turkiye	Orginal article	Nursing	Image - Sound	1	DALL-E3, Image-Gen and D-ID program	Yes	Nursing care / total hip arthroplasty	No	No	Yes	188 third-year nursing students (94 AI group, 94 control)	Yes (satisfaction evaluation form, knowledge test, f ive priority nursing diagnoses) - LIKERT	AI group superior in case management and diagnosis accuracy.	No
Arain et al.	Saudi Arabia	Orginal Article	Medicine	Text	1	Gemini Advanced	No	Diabetus mellitus	Yes	Yes	No		Yes (Assessed by four independent faculty members using a structured questionnaire with a 3-point Likert scale.) - LIKERT	LMM case better alignment with objectives and for providing appropriate triggers; however, weaker in realism and level-appropriateness.	No
Liu et al,	China	Orginal Article	Medicine	Text	8289	Qwen, Baichuan, Erine	No	hypertension, diabetes, hepatitis B	Suplementa ry material	No	No		No	No	Yes (The cases have produced biased scenarios at various levels based on education, gender, income, and insurance status.))
Xie et al,	China	Orginal Article	Medicine	Image	150	Stable Diffusion 1.5	No	Infectious keratitis	No	No	Yes	97 fourth-year medical students	Yes (30-question test featuring real patient images was formally administered both before and after the interventions to all students, (three groups: taught with real cases; taught with AI-generated images and taught with real medical images). - MODEL BASED	The teaching effectiveness of AI-generated images was comparable to that of real medical images	No