INPLASY

The Impact of Generative Artificial Intelligence on Clinical Skills and Knowledge Acquisition in Medical Undergraduates: A Systematic Review and Meta-Analysis

INPLASY2025120075
doi: 10.37766/inplasy2025.12.0075
Received: 20 December 2025
Published: 20 December 2025

Xie, GL; Liao, JL; Tang, XX; Fu, H; Liu, D; Deng, L; Wang, T; Hong, X.

Corresponding author:
Guanli Xie

mianmian186@sina.cn

Author Affiliation:
Yunnan University of Chinese Medicine.

ADMINISTRATIVE INFORMATION

Support - This study was funded by the Yunnan Province High-level Chinese medicine talents project (2024), Yunnan Province's Healthcare Professionals of the "Xingdian Talent Support Program" (XDYC-YLWS-2024-0018), Yunnan Province's Youth Talent Project of "Xingdian Talent Support Program" (XDYC-QNRC-2024-523), and Yunnan Province doctoral student research team of acupuncture, massage for treating brain diseases (10170101868), and Yunnan Province Clinical Excellent Talent Training Program for Traditional Chinese Medicine Practitioners (YuCai[2025] No. 112).

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY2025120075

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 20 December 2025 and was last updated on 20 December 2025.

INTRODUCTION

Review question / Objective To systematically evaluate the application effects of generative AI in the education and teaching of medical undergraduates (including nursing and medical majors), including its impact on outcome indicators such as students' theoretical knowledge mastery, practical skill operation, clinical thinking ability, and learning satisfaction.

Condition being studied With the rapid advancement of artificial intelligence technology, Generative Artificial Intelligence (GAI), such as ChatGPT, GPT-4, and Med-PaLM, has gradually penetrated the field of medical education due to its powerful capabilities in natural language processing, knowledge generation, and interactive feedback. Medical undergraduates, as the core group in medical talent cultivation, their educational quality is directly related to the sustainable development of the healthcare industry. Undergraduate education in nursing and medical majors (e.g., clinical medicine, preventive medicine, laboratory medicine) emphasizes both the systematicness of theoretical knowledge and the standardization of practical skills and the cultivation of clinical thinking.
Currently, several studies have explored the application effects of generative AI in medical undergraduates' curriculum learning, skill training,

assessment, and evaluation. However, the research results are heterogeneous: some studies have shown that generative AI can improve students' learning efficiency, knowledge mastery, and clinical problem-solving abilities, while others have pointed out its limitations in practical skill training and ethical cognition guidance. Due to differences in sample size, intervention protocols, and evaluation indicators among individual studies, it is difficult to comprehensively and objectively reflect the overall impact of generative AI on the education and teaching of medical undergraduates. Therefore, integrating evidence from relevant controlled studies through Meta-analysis can provide an evidence-based basis for medical educators to optimize teaching plans and rationally apply generative AI.

METHODS

Participant or population Full-time medical undergraduates; majors include nursing and medical disciplines.If multiple populations were included in the original study, only those studies where the proportion of the target population exceeded 80% would be included.

Intervention The intervention group adopts generative AI for teaching intervention, including but not limited to the following forms: AI-assisted course teaching (e.g., intelligent courseware generation, personalized knowledge point push), AI-guided autonomous learning (e.g., intelligent Q&A, learning plan customization), AI-simulated practical training (e.g., virtual patient consultation, nursing operation simulation feedback), AI-assisted assessment and evaluation (e.g., automatic correction of objective questions, subjective question scoring suggestions, learning effect analysis); types of generative AI tools are not limited (e.g., general-purpose AI: ChatGPT, ERNIE Bot; medical-specific AI: Med-PaLM, ChatMed, etc.)

Comparator Control intervention methods are not limited, including traditional teaching models (e.g., offline classroom lectures, blackboard teaching, conventional practical tutoring), non-generative AI-assisted teaching.

Study designs to be included Randomized Controlled Trials (RCTs) and Non-Randomized Controlled Trials (NRCTs).

Eligibility criteria Chinese and English literature.

Information sources Medline (via PubMed), Embase (via Embase.com), CINAHL (via EBSCOhost), APA PsycArticle (via EBSCOhost), APA PsycInfo (via EBSCOhost), Educational Resources Information Center (ERIC) (via EBSCOhost), Scopus, China National Knowledge Infrastructure (CNKI), WanFang Database, China Science and Technology Journal Database (VIP), and Chinese Biomedical Literature Service System (SinoMed).

Main outcome(s) Basic knowledge, clinical skills, academic writing scores.

Quality assessment / Risk of bias analysis The Medical Education Research Study Quality Instrument (MERSQI)and the Risk of Bias (ROB2) tool were used to evaluate the quality of the studies.

Strategy of data synthesis Heterogeneity among studies were analyzed using the I^2 test. I^2 indicates low heterogeneity, 25%-50% indicates moderate heterogeneity, and >50% indicates high heterogeneity. If $I^2 \leq 50\%$ and $P \geq 0.10$, a fixed-effects model were used to pool effect sizes; if $I^2 >$ 50% or P .10, a random-effects model were adopted, and sources of heterogeneity were analyzed.Continuous outcome indicators (e.g., exam scores, satisfaction scores): Weighted Mean Difference (WMD) or Standardized Mean Difference (SMD) with 95% Confidence Interval (CI) were used.

Subgroup analysis Subgroup analyses were performed according to major type (nursing major/ medical major).

Sensitivity analysis Sensitivity analysis were conducted by sequentially excluding each individual study and re-pooling effect sizes to test the stability of the results.

Country(ies) involved China.

Keywords Generative artificial intelligence, Undergraduate medical education, Digital education, Meta-analysis.

Dissemination plans Generative artificial intelligence, Undergraduate medical education, Digital education, Meta-analysis.

Contributions of each author
Author 1 - Guanli Xie.
Email: mianmian186@sina.cn
Author 2 - Jianglong Liao.
Email: liao13988902077@126.com
Author 3 - Xiaoxia Tang.
Email: 405269710@qq.com

Author 4 - Han Fu.
Email: 540570411@qq.com
Author 5 - Duo Liu.
Email: 48911297@qq.com
Author 6 - Li Deng.
Email: 479404017@qq.com
Author 7 - Tao Wang.
Email: taowang0871@126.com
Author 8 - Xin Hong.
Email: xinghong0877@163.com