INPLASY

Effectiveness of Generative Artificial Intelligence-Based Teaching Versus Traditional Teaching Methods in Medical Education: A Meta-Analysis of Randomized Controlled Trials

INPLASY202510006

doi: 10.37766/inplasy2025.1.0006

Received: 2 January 2025

Published: 2 January 2025

Corresponding author:
Xiaoqin Jiang

1598862657jxq@scu.edu.cn

Author Affiliation:
Department of Anesthesiology, West China Second University Hospital, Sichuan University.

Li, J; Yin, KY; Chen, DX; Jiang, XQ.

ADMINISTRATIVE INFORMATION

Support - No funding support.

Review Stage at time of this submission - Piloting of the study selection process

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202510006

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 6 January 2025 and was last updated on 6 January 2025.

INTRODUCTION

Review question / Objective We included RCT studies in the analysis to determine whether medical students who learn with generative artificial intelligence demonstrate an improvement in their theoretical knowledge or clinical skills scores compared to traditional learning methods.

Condition being studied The swift advancement of artificial intelligence (AI) technology in recent years has affected a variety of sectors, including medical education and healthcare. Generative artificial intelligence (GAI) is a deep learning-based AI that uses algorithms trained on large datasets to produce new content in a variety of formats (text, audio, video, etc.).

GAI demonstrates significant potential in the field of education. For example, learning can be more targeted and efficient by tailoring study plans to the unique needs of each student. And also, GAI offers immediate, constructive feedback and assessments and enhances students' communication and practical skills through simulated dialogues.

Teacher-led lectures or online learning platforms are the mainstays of traditional teaching methods in modern medical education. Although these methods encourage the acquisition of knowledge, students frequently fall short in generating motivation for self-directed learning and facilitating opportunities for repeated practice. GAI might present a novel paradigm for inventive and adaptable teaching. However, issues with plagiarism, reliance, and the veracity of the content generated on GAI have been concerned as well.

There are currently no high-caliber meta-analyses to support assertions that GAI performs better than traditional teaching methods. Thus, this study aims to compare the effectiveness of GAI-based teaching and traditional teaching methods in medical education.

METHODS

Search strategy

PubMed:

("Artificial Intelligence"[Mesh] OR "Artificial Intelligence"[Title/Abstract] OR "Generative Artificial Intelligence"[Title/Abstract] OR "Generative AI"[Title/Abstract] OR "Artificial Intelligence Chatbot"[Title/Abstract] OR "AI Chatbot"[Title/Abstract] OR "Large Language Model"[Title/Abstract] OR "Generative Language Model"[Title/Abstract] OR "ChatGPT"[Title/Abstract]) AND ("Education, Medical"[Mesh] OR "Medical Education"[Title/Abstract] OR "Teaching"[Title/Abstract] OR "Teaching Method*"[Title/Abstract] OR "Medical Learning"[Title/Abstract]) AND ("randomized controlled trial"[Publication Type] OR "random*"[All Fields] OR "controll*"[All Fields] OR "trial"[All Fields]) AND (y_10[Filter])

Embase:

('generative artificial intelligence'/exp OR 'generative artificial intelligence' OR 'generative ai'/exp OR 'generative ai' OR 'artificial intelligence chatbot'/exp OR 'artificial intelligence chatbot' OR 'ai chatbot'/exp OR 'ai chatbot' OR 'large language model'/exp OR 'large language model' OR 'generative language model' OR 'chatgpt'/exp OR 'chatgpt') AND ('medical education'/exp OR 'medical education' OR 'teaching'/exp OR 'teaching' OR 'teaching method*' OR 'learning'/exp OR 'learning') AND ('randomized controlled trial'/exp OR 'randomized controlled trial' OR 'random*':ab,ti OR 'controll*':ab,ti OR 'trial':ab,ti) AND [2014-2024]/py

Cochrane Library:

#1 MeSH descriptor: [Artificial Intelligence] explode all trees

#2 ("Generative Artificial Intelligence" OR "Generative AI" OR "Artificial Intelligence Chatbot" OR "AI Chatbot" OR "Large Language Model" OR "Generative Language Model" OR "ChatGPT"):ti,ab,kw

#3 MeSH descriptor: [Education, Medical] explode all trees

#4 ("Education" OR "Teaching" OR "Teaching Method" OR "Medical Learning"):ti,ab,kw

#5 (#1 OR #2) AND (#3 OR #4) with Publication Year from 2014 to 2024, in Trials.

Participant or population Medical students.

Intervention Learn with generative artificial intelligence tools.

Comparator Learn with traditional teaching methods, including traditional lectures, textbooks, online searches, and other non-AI methods.

Study designs to be included Randomized Controlled Trials.

Eligibility criteria Inclusion Criteria: RCTs comparing GAI and traditional teaching methods in medical education.

Exclusion Criteria: non-RCT articles, non-GAI studies, studies unrelated to medical students, and articles for which the full text is unavailable.

Information sources Search electronic databases, including PubMed, EMBASE, and Cochrane Library, and search the references of the included articles.

Main outcome(s) The primary outcomes include scores in theoretical knowledge or practical skills after learning with GAI.

Additional outcome(s) The secondary outcomes include students' perceptions of GAI, self-learning interest, learning status, and so on.

Quality assessment / Risk of bias analysis The quality of each outcome was rated using the GRADE group, categorized as very low, low, moderate, or high quality.

Each RCT was evaluated for risk of bias using the Cochrane collaboration risk of bias tool, classifying studies as high, low, or unclear risk.

Strategy of data synthesis Meta-analyses were conducted using Review Manager (RevMan for Windows, Version 5.4). Risk ratios (RR) with 95% confidence intervals (CI) were calculated for dichotomous variables. Standardized mean differences (SMD) with 95% CI were calculated for continuous variables. Variance was assessed using the chi-square test. A p-value < 0.05 was considered statistically significant.

Subgroup analysis Subgroup analysis was planned to explore potential differences in primary outcomes based on the duration of education and types of GAI-based teaching. Furthermore, both short- and long-term knowledge and skill retention were investigated.

Sensitivity analysis Perform sensitivity analysis through different subgroups and model construction.

Country(ies) involved China.

Keywords generative artificial intelligence; education; medical students.

Contributions of each author

Author 1 - Juan Li.
Author 2 - Kaiyu Yin.
Email: kaiyuyin@126.com
Author 3 - Dongxu Chen.
Author 4 - Xiaoqin Jiang