The Evolving Role of ChatGPT (Chat-Generative Pre-Trained Transformer) in General Surgery: A Systematic Review

Raquel Nogueira MD[1] ; Carlos A. Balthazar da Silveira MD[2] ; Ana Caroline D. Rasador MD[2]; Masashi Takeuchi MD3; Yuko Kitagawa MD[3];
Flavio Malcher MD4; B Todd Heniford MD5;Diego L. Lima MD MSc[1]
Montefiore Medical Center[1], Dignity Health St. Joseph's Hospital[2], Keio University[3], Lenox Hill Hospital, Northwell[4], Carolinas Medical Center[5]

SAGES 2025

Montefiore THE UNIVERSITY HOSPITAL

Introduction

- Chat Generative Pre-Trained Transformer (ChatGPT) emerged in 2020 as one
- of the pioneering accessible artificial intelligence (AI) models online.
- Despite gaining worldwide recognition, initial medical literature on ChatGPT surfaced in 2022, exploring its capabilities in drug and disease generation.
- As no systematic review has examined
- ChatGPT's role in general surgery, we performed a systematic review to explore ChatGPT'S potential in general surgery.

Methods

- PubMed/MEDLINE, EMBASE, and Cochrane, Scielo e Lilacs databases were searched without language or date restriction focusing on studies that explored ChatGPT possible applications in educational, research, and clinical practice was performed.

Main Findings

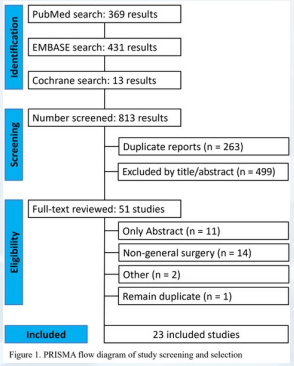


Figure 1. PRISMA flow diagram of study screening and selection

Table 1, 2 and 3. Educational, Clinical practice/decision-making and Research applications of ChatGPT.

- ChatGPT's diverse applications in surgical contexts: 6 demonstrated its capability in addressing common questions about surgical diseases, 7 highlighted its utility in clinical practice, 11 explored its educational potential, and 5 examined its research functions.
- ChatGPT exhibited proficiency in providing anatomical explanations and answering open-ended questions, achieving an accuracy of approximately 87% for colorectal queries but varied performance for appendicitis questions.
- Assessments of its performance in board exam-style questions yielded accuracy rates ranging from 48% to 66% for open-ended questions and 68% to 76.4% for multiple-choice questions.
- In addressing patient queries, ChatGPT's responses were generally rated positively, particularly in bariatric, transplant, and pancreatic surgery domains. Additionally, ChatGPT displayed enhanced clarity and 10 comprehensiveness compared to traditional medical literature and demonstrated efficiency in generating reliable questions with minimal revisions.

Conclusion

- ChatGPT shows significant potential to be explored in the realms of surgical
- education, clinical practice, and research. However, it is not yet an independent tool from
- human evaluation, as all its potential uses must be assessed by an experienced surgeon.

References