

Exploring the role of an artificial intelligence chatbot on 38 appendicitis management: An experimental study on ChatGPT



Authors: Dylan Gracias, Adrian Siu, Ishith Seth, Dilshad Dooreemeah, Angus Lee

Introduction

- ❖ Appendicitis is a common surgical condition that requires urgent medical attention
- ❖ ChatGPT is a deep learning language model that has gained widespread recognition for its versatility and potential applications in clinical medicine
- ❖ As AI and machine learning continue to progress, these systems could significantly enhance the clinician's and patient's access to reliable, high quality, prompt medical information

Aims

- ❖ To assess the precision and comprehensiveness of answers generated by ChatGPT regarding appendicitis management, compared to clinical guidelines and evidence-based literature
- ❖ Assess the quality of references provided by ChatGPT

Methods

- ❖ ChatGPT was presented with six questions related to appendicitis
- ❖ Responses and references were analysed for accuracy, comprehensiveness, coherence, appropriateness, and quality against evidence-based literature
- ❖ Further evaluation was performed using a five point Likert scale

Question	Key Positive Points	Key Negative Points	References
1. <i>What is the cause of appendicitis?</i>	<ul style="list-style-type: none">▪ Prefaced that the exact mechanism is not fully understood▪ Identified obstruction as a theory, including fecaliths, lymphoid hyperplasia and tumours contributing to this	<ul style="list-style-type: none">▪ No elaboration on obstruction leading to inflammation▪ Did not identify any other pathophysiological theories▪ Did not identify environmental, infectious, or ethnic factors which may contribute▪ Repetitive rather than diverse and specific	<ul style="list-style-type: none">▪ 2 existent, relevant publications▪ 3 existent, less relevant publications which deviated from the question
2. <i>What are the preventative measures for appendicitis?</i>	<ul style="list-style-type: none">▪ Included hygiene precautions	<ul style="list-style-type: none">▪ Failed to acknowledge the limited evidence on the prevention of appendicitis▪ Hygiene precautions was non-descriptive▪ Mentioned techniques found throughout the internet but not validated by evidence-based research (eg: exercise, stress reduction)	<ul style="list-style-type: none">▪ 4 non-existent publications▪ 1 existent, but irrelevant publication
3. <i>What is the most appropriate management for appendicitis?</i>	<ul style="list-style-type: none">▪ Included appendectomy and antibiotic treatment with a description of each	<ul style="list-style-type: none">▪ Provided limited information on the specific indications or considerations for each treatment option▪ No data on success rates, complication rates, recurrence rates, negative appendectomy rates, length of hospital stay and economic implications to compare these two treatment options▪ Lacked depth	<ul style="list-style-type: none">▪ 2 existent, high quality, relevant studies▪ 3 existent, but irrelevant studies▪ Missed key randomised trial, systematic review, and meta-analysis studies
4. <i>Compare the efficacy between surgical versus antibiotics for appendicitis management and what is superior?</i>	<ul style="list-style-type: none">▪ More comprehensive▪ Specified the severity of the disease contributing to the decision, amongst other factors▪ Discussed recurrence rate associated with antibiotic therapy▪ Downsides of antibiotic therapy were explored	<ul style="list-style-type: none">▪ No quantitative data on success or recurrence rates▪ Downsides of surgical management not explored	<ul style="list-style-type: none">▪ 5 existent, high quality, relevant publications
5. <i>Provide future recommendations for advancement in management of appendicitis and what further research is needed?</i>	<ul style="list-style-type: none">▪ Suggested further evaluation of medical imaging in the diagnostic process	<ul style="list-style-type: none">▪ Largely inconsistent with current future directions in this field (eg: robotic assisted appendectomy, intraoperative imaging, immunomodulation)	<ul style="list-style-type: none">▪ 2 existent, valid publications▪ 2 existent, but irrelevant publications▪ 1 non-existent publication
6. <i>What is the incidence of appendicitis in pregnancy?</i>	<ul style="list-style-type: none">▪ Accurate in identifying that pregnancy is a protective factor▪ Emphasised the importance of prompt diagnosis and treatment to prevent maternal and foetal complications	<ul style="list-style-type: none">▪ Scattered statistics▪ No comparison of incidence to the general population nor the non-gravid female population	<ul style="list-style-type: none">▪ 5 non-existent publications

Discussion

- ❖ Breadth of knowledge, fast processing speed, and ability to present the information succinctly
- ❖ Inconsistent in its accuracy, with answers frequently lacking in depth and specificity
- ❖ Commonly referenced fabricated or irrelevant publications, despite the existence of pertinent and real sources
- ❖ Integration of such technology could promote improved healthcare in resource limited settings
- ❖ Ongoing legal and ethical issues of the AI, including patient privacy, plagiarism of resources and spread of misinformation

Conclusion

- ❖ Demonstrates great potential but none the less numerous pitfalls which question its use in clinical practice
- ❖ Necessitates further refinement before integration into the clinical settings
- ❖ Clinicians and all users must exercise vigilance when utilising ChatGPT for medical information purposes

