

Customizing a Large Language Model (LLM) to Generate SAGES Guideline Recommendations for the Surgical Management of GERD.

Bright Huo[1], Elisa Calabrese,[2] Sunjay Kumar,[3] Romeo C. Ignacio,[4] Rodolfo Oviedo,[5-7] Imran Hassan,[8] Bethany Slater,[9] Danielle S. Walsh,[10] Wesley Vosburg.[11]

[1]McMaster University, Canada; [2]University of California South California, USA; [3]Thomas Jefferson University Hospital, USA; [4]University of California, USA; [5]Nacogdoches Center for Metabolic and Weight Loss Surgery, USA; [6]University of Houston Tilman J. Fertitta Family College of Medicine, USA; [7]Sam Houston State University College of Osteopathic Medicine, USA; [8]University of Iowa, USA; [9]University of Chicago, USA; [10]University of Kentucky, USA; [11]Harvard University, USA

McMaster University Surgery

Introduction

LLM-linked chatbots may be an efficient source of clinical recommendations for healthcare providers and patients. However, LLM development is resource-intensive.

Objectives

To evaluate the ability of a customized ChatGPT-4 model to provide guidance for the surgical management of GERD to both surgeons and patients.

Methods

Forty-five patient cases created based on SAGES guidelines for the surgical treatment of GERD.

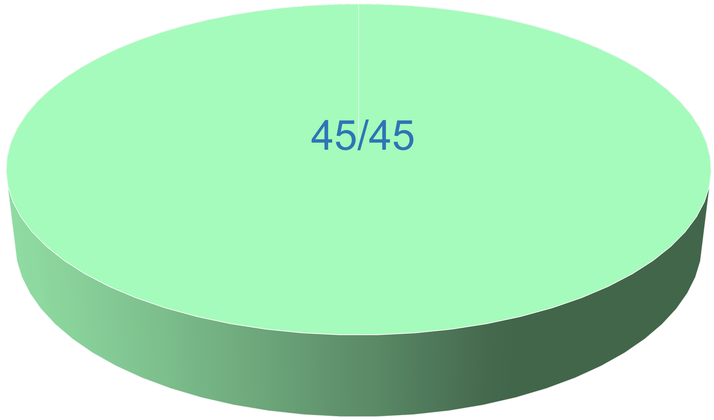
A customized ChatGPT-4 model developed to tailor its responses by uploading the SAGES GERD guideline and using instructional and test prompts to guide its use based on recommendations, creating the SAGES GERD Tool for Surgery (SGTS).

ChatGPT-4 & SGTS queried Dec 3rd, 2023 with standardized prompts for advice on the surgical management of GERD.

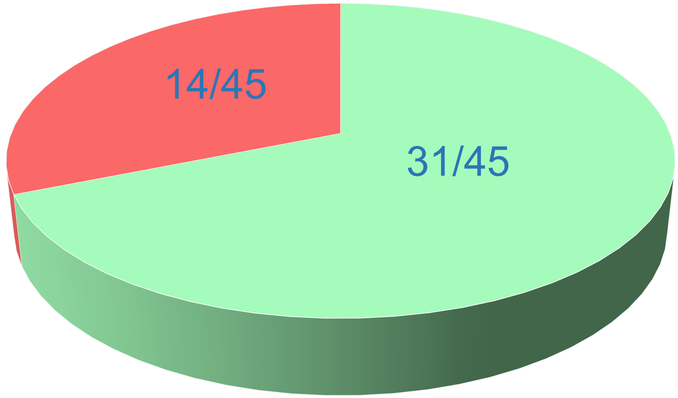
Evaluation based on alignment with guideline recommendations.

Results

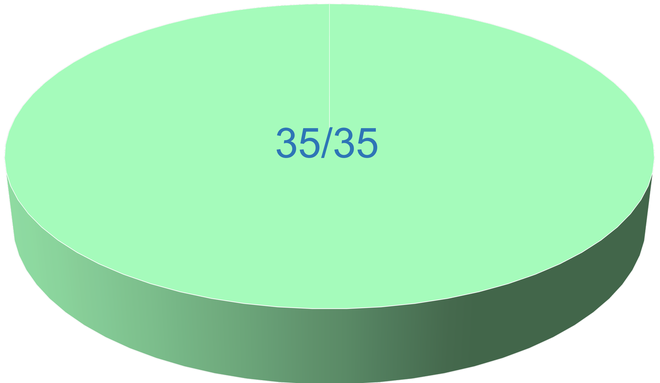
SGST
Surgeon Advice



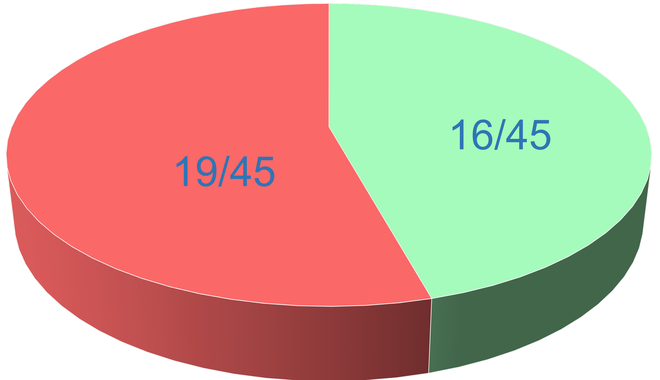
ChatGPT-4
Surgeon Advice



SGST
Patient Advice



ChatGPT-4
Patient Advice



Results

SGST answered all questions correctly.

ChatGPT-4 correct 68.9% (surgeon advice) & 35.6% (patient advice)

Conclusions

ChatGPT-4.0 can be customized to overcome limitations with its training dataset to provide recommendations for the surgical management of GERD with reliable accuracy and consistency.

Clinicians, patients, policymakers, and hospital managers should take note that this methodology may be applied to support surgical decision-making.

Future prospective data is needed to assess its effectiveness in a pragmatic clinical environment.