Tilmelding af Foredrag

Foredragets titel

Validity of the large language model ChatGPT (GPT4) as a patient information source in otolaryngology by a variety of doctors in a tertiary otorhinolaryngology department

Forfatter(e)

Jacob P. S. Nielsen, Christian von Buchwald and Christian Grønhøj

Afdeling/praksis

Alle med samme afdeling: Afdeling for Øre-Næse-Halskirurgi og Audiologi, Rigshospitalet

Uddannelsesniveau

Stud.Med.

Introduktion

A high number of patients seek health information online, and large language models (LLMs) may produce a rising amount of it. This study aims to evaluate the performance regarding health information provided by ChatGPT, a LLM developed by OpenAI, focusing on its utility as a source for otolaryngology-related patient information.

Materiale/metode

A variety of doctors from a tertiary otorhinolaryngology department used a Likert scale to assess the chatbot's responses in terms of accuracy, relevance, and depth. The responses were also evaluated by ChatGPT.

Resultater

The composite mean of the three categories was 3.41, with the highest performance noted in the relevance category (mean = 3.71) when evaluated by the respondents. The accuracy and depth categories yielded mean scores of 3.51 and 3.00, respectively. All the categories were rated as 5 when evaluated by ChatGPT.

Diskussion

In this study ChatGPT was evaluated for the relevance, accuarcy and depth of the provided medical information. It scored high in relevance (3.71) but lower in depth (3.00), struggling with complex diagnoses. Limitations include knowledge cut-off in 2021 and potential training data biases. Despite this, ChatGPT shows promise in healthcare. The study highlights the need for improvements in depth and bias management, however also suggesting a big potential role for chatbots in future healthcare.

Forfatters fulde navn

Jacob Pohl Stangerup Nielsen

Forfatters email

jacob.pohl.stangerup.nielsen.03@regionh.dk