ChatGPT based Layman Explanation and Translation of Radiology Reports via ChatGPT: Is it any good?

Emiliano Garza Frias MD, Parisa Kaviani MD, Giridhar Dasegowda MBBS, Jinghu Hu MD PhD, Lina Karout MD, Roshan Fahimi MD, Mannudeep Kalra MD, Bernardo Bizzo MD PhD, Keith Dreyer DO PhD FACR
Massachusetts General Hospital – Mass General Brigham

Purpose

Large Language Models (LMMs) might represent the future of internet interactions and queries. We assessed the performance of ChatGPT for explaining and translating radiology reports in layman terms for common public use.

Materials and methods

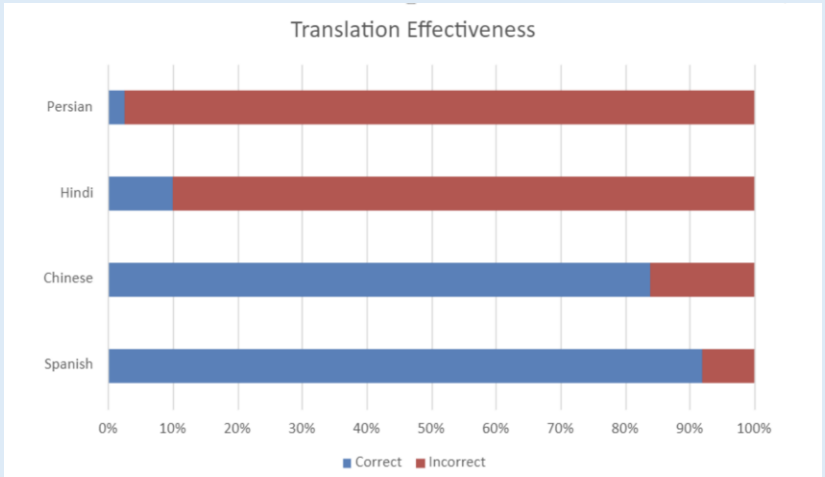
The study included 120 radiology reports from 120 adult patients belonging to chest CT, abdomen CT, brain MR, and chest radiography (30 for each imaging modality). First, we searched and exported completely de-identified radiology reports from our radiology report search engine. Next, we copied the impression section of the reports and pasted each radiology report separately under the question headers "Today I got my radiology report and I do not understand what the radiologist wrote on the impression. Could you explain it to me and tell me which are the next steps in my management?". ChatGPT outputs were reviewed by a general physician. In addition, we requested ChatGPT to translate and explain radiology findings in the impression sections separately in Spanish, Chinese, Hindi, and Persian. Data were reviewed by native speaking physicians to verify accuracy and understandability of ChatGPT outputs.

Results

ChatGPT correctly explained an average of 4 medical terms per impression section in 120 radiology reports.

About 94% of the medical terms included in this section were correctly explained in layman terms. Vague recommendations about treatment, lab testing and follow up differed between imaging modalities and findings.

The translation capabilities of Chat GPT varied tremendously across different languages (Spanish 92%; Chinese 84%; Hindi 10%; Persian 2.5%).



Conclusions

ChatGPT performs admirably in explaining radiology impressions in layperson English and Spanish but falls short when it comes to translation to Hindi and Persian. Most ChatGPT recommendations of further action based on radiology impression are vague or incorrect. Large Language Models (LMMs) are showing rapid advancements in general interpretation and querying of broad domains of text. With the newness of this technology, we expect these LMMs to rapidly improve in this domain.

A perspective for the future

This study was performed at the beginning of 2023. Since then, Chat GPT has undergone 6 updates with introduction of ChatGPT 3.5 and 4.0, the ability to interact with third-party applications and its connection to the World Wide Web.

New preliminary results with ChatGPT 4.0 demonstrate superior translation capabilities. Further evaluation will help evaluate whether these resources are safe for public consumption.