LETTERS TO THE EDITOR

CHATGPT APPLIED TO SOLVE VIRTUAL MEDICAL SCENARIOS

CHATGPT APLICADO PARA RESOLVER ESCENARIOS MÉDICOS VIRTUALES

Gonzalo Vidangos-Paredes [iD][1,a], Elizabeth Valeria Rijalba-Monsefú [iD][1,a]

LETTERS TO THE EDITOR

Dear. Editor:

Currently, patients tend to seek information about their illnesses on the internet. While many of these sources are reliable, others are not. ChatGPT, as an artificial intelligence (AI) tool, has the potential to discern between these sources and provide more accurate answers. In recent years, the use of AI in the medical field has significantly increased. Numerous studies have evaluated ChatGPT's ability to answer medical questions, ranging from simple to complex, similar to those used in medical licensing exams.

For instance, a study demonstrated that the ChatGPT-4 version successfully surpassed the passing threshold of the National Medical Examination in Japan, whereas the previous version, ChatGPT-3.5, did not[1]. However, in China, a similar study resulted in the AI failing the exam[2]. In the United States (US), ChatGPT was evaluated in the United States Medical Licensing Examination (USMLE), using two question banks from Step-1 and Step-2, achieving satisfactory results[3].

In our country, a study was conducted using both ChatGPT-3.5 and ChatGPT-4 to answer the National Medical Examination (ENAM, by its Spanish acronym), and both versions passed the exam. Additionally, ChatGPT's accuracy exceeded that of the students who were evaluated, with scores of 86%, 77%, and 55%, respectively [4]. In this context, a study was conducted in January of this year to evaluate the effectiveness of ChatGPT-3.5 in solving basic virtual medical scenarios, specifically on Chronic Obstructive Pulmonary Disease (COPD) and multimorbidity. The Virtual Patients Scenarios App platform, developed by the Medical Physics and Digital Innovation Lab of the Faculty of Medicine of the Aristotle University of Thessaloniki, Greece, was used. The "Symptom Management Scenarios" section was accessed, selecting "Symptom Management: COPD" and "Symptom Management: Multimorbidity." The questionnaires included six and nine dynamic questions, respectively, based on patient simulations, each taking approximately five minutes.

First, the questions were answered manually, and then ChatGPT was asked to answer the same questions. The responses were later tabulated and outlined using Microsoft® Excel for Mac version 16.78.3.

Journal home page: http://revistas.urp.edu.pe/index.php/RFMH

For the COPD scenario, ChatGPT answered one out of six questions incorrectly, achieving an accuracy rate of 83.33%. For the multimorbidity scenario, three out of nine answers were incorrect, resulting in 66.67%.

Overall, ChatGPT achieved a 73.33% accuracy rate across both scenarios, answering eleven out of fifteen questions correctly (Figure 1).

COPD Scenario

Multimorbidity Scenario

1 (16,7%)

5 (83,3%)

3 (33,3%)

6 (66,7%)

● CORRECT ● INCORRECT

● CORRECT ● INCORRECT

Both Scenarios

11

4

0   2   4   6   8   10   12

■ CORRECT ■ INCORRECT
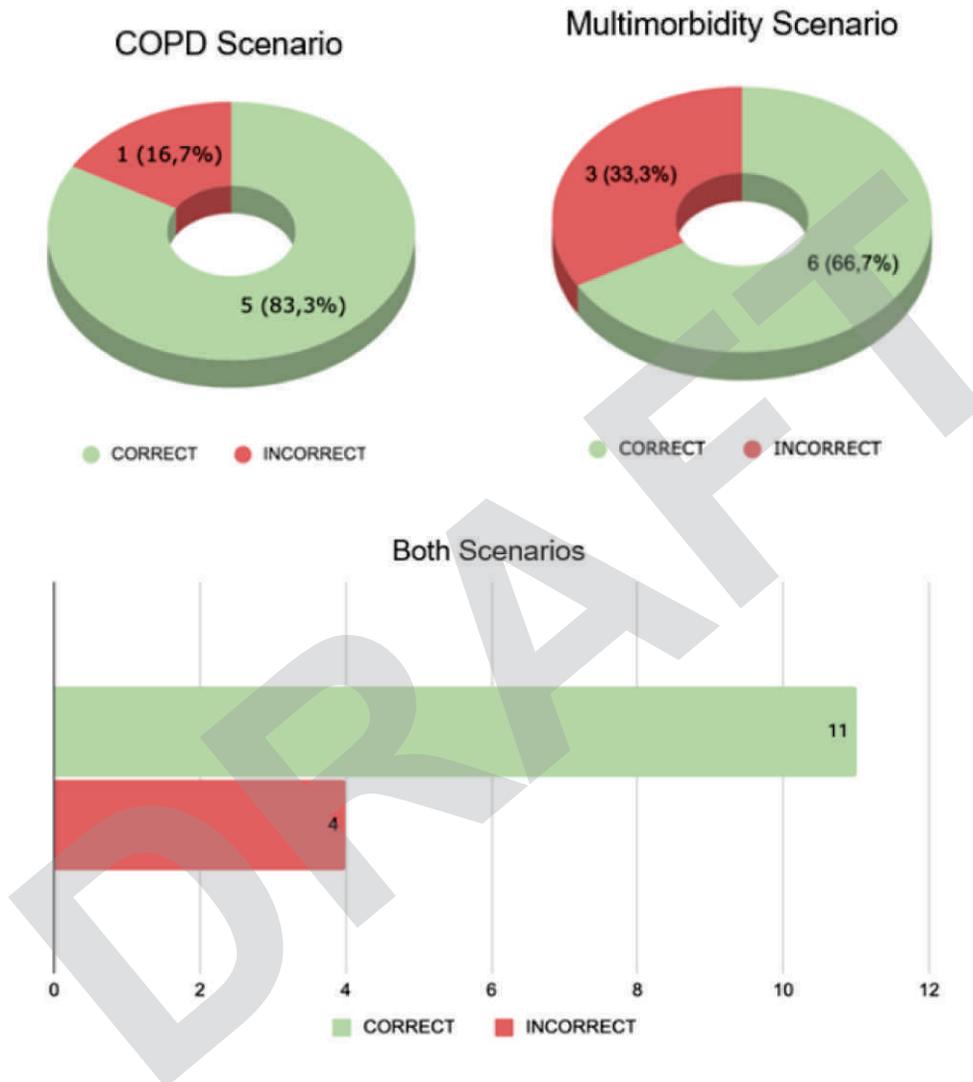
Figure 1A. ChatGPT results in specific scenarios.

Figure 1B. ChatGPT results in both scenarios.

While AI can provide general and relevant information, it should not be considered a substitute for the clinical judgment of healthcare professionals. There are still significant gaps, such as the lack of personalization, the risk of incorrect information, and the ethical and liability implications. In this study, the error margin was 26.67%, raising concerns about trust in the application.

Although this pilot study included only 15 questions, it can be compared to the work of Soto-Chávez et al. [5], an analytical observational cross-sectional study that evaluated 12 questions selected by internal medicine specialists on five chronic diseases (diabetes, heart failure, chronic kidney disease, rheumatoid arthritis, and systemic lupus erythematosus). That study found that 71.67% of the responses generated

by ChatGPT were rated as "good," and none were considered "completely incorrect," with higher accuracy in diabetes and rheumatoid arthritis. In conclusion, various studies have evaluated AI in medical licensing exams in different countries. However, there are limited studies that investigate its ability to answer questions on specific diseases. As observed in this pilot study, ChatGPT can tackle specific medical scenarios, providing general information and answers based on the knowledge it has acquired during its training. Nevertheless, it is crucial to remember that ChatGPT is not a medical professional and has limitations. It should not be regarded as a replacement for consultation with a qualified medical expert, let alone as a means for self-diagnosis.

This study could serve as an inspiration for future research comparing various AI tools and their ability to address different diseases. Moreover, this type of study does not demand significant financial costs or considerable time, as mostly free-access virtual tools can be used.

LETTERS TO THE EDITOR

Correspondence: *Gonzalo Vidangos-Paredes.*
Address: *Av. Monterrico Sur 120, 303, Santiago de Surco, Lima, Perú.*
Telephone number: *(+51) 950 445 531*
Email: *gonzalovidangos@me.com*

REFERENCES

1.Yanagita Y, Yokokawa D, Uchida S, Tawara J, Ikusaka M. Accuracy of ChatGPT on Medical Questions in the National Medical Licensing Examination in Japan: Evaluation Study. JMIR Form Res. 2023;7. doi: 10.2196/48023.

2.Wang X, Gong Z, Wang G, Jia J, Xu Y, Zhao J, et al. ChatGPT Performs on the Chinese National Medical Licensing Examination. J Med Syst. 2023;47(1):86. doi: 10.1007/s10916-023-01961-0.

3.Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ. 2023;9. doi: 10.2196/45312.

4.Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez SF, Cruz-Galán JD, Gutiérrez-Arratia JD, Torres BGQ, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: Cross-Sectional Study. JMIR Med Educ. 2023;9(1). doi: 10.2196/48039.

5.Soto-Chávez MJ, Bustos MM, Fernández-Ávila DG, Muñoz OM. Evaluation of information provided to patients by ChatGPT about chronic diseases in Spanish language. Digit Health. 2024;10:1-7. doi: 10.1177/20552076231224603.