SOHO ANNUAL MEETING 2024

Evaluating the Efficacy and Accuracy of ChatGPT in Providing Chemotherapy Education to Lymphoma Patients

Ahmad Basharat MD[1] , Ammad Naeem MD[2], Yuanli Lei MD[3], Aqsa Mumtaz MD[4], Iqra Naeem MD[5], Ahmad Zahid MD[6], Moazzam Shahzad MD[7]

[1] Marshfield Clinic Health System, Marshfield, WI, USA; [2] WVU Medicine, Thomson Medical Center, WV, USA; [3] Marshfield Clinic Health System, Marshfield, WI, USA; [4] Montefiore St. Luke's Cornwall, Newburgh, NY, USA; [5] Marshfield Clinic Health System, Marshfield, WI, USA; [6] Marshfield Clinic Health System, Marshfield, WI, USA; [7] Moffitt Cancer Center, University of South Florida, Tampa, USA

MOFFITT CANCER CENTER

INTRODUCTION

The Chat Generative Pre-Trained Transformer (ChatGPT) signifies a major advancement in artificial intelligence (AI) chatbot technology. Despite its promising capabilities, concerns about its reliability and accuracy remain. This study aims to evaluate ChatGPT's responses to frequently asked questions about chemotherapy education in Lymphoma Patients

METHOD

Experienced oncologists and chemotherapy nurses who regularly provide patient education were invited to evaluate ChatGPT-generated answers to 8 frequently asked chemotherapy questions. These answers were compared to those sourced from hospital websites. The evaluation criteria included ease of understanding, scientific accuracy, and overall satisfaction. Additionally, evaluators attempted to identify whether each response was AI-generated

RESULTS

- Among the 10 respondents, 3 were female oncologists (30%) and 7 were male.
- Regarding familiarity with ChatGPT, 2 respondents had never used it, while 2 reported regular use. The average ChatGPT usage level among participants was 2.75 on a scale from 1 to 5.
- Out of 31 answers provided, 8 were generated by ChatGPT. Evaluators correctly identified AI-generated responses 54% of the time, with a sensitivity of 39% and specificity of 60%. On average, evaluators correctly identified 17 out of 31 AI-generated responses.
- Factors such as gender, prior experience with chemotherapy education, and the frequency of ChatGPT usage did not significantly impact the detection rate of AI-generated responses. The scientific accuracy of the 8 AI-generated answers averaged 3.81 on a scale from 1 to 5, slightly higher than the average score of 3.78 for non-AI-generated answers. AI-generated answers received an overall score of 4.11 for ease of understanding, surpassing the 4.02 score for non-AI-generated answers.
- Regarding overall satisfaction, AI-generated responses scored 3.87 compared to 3.71 for non-AI-generated answers. These differences were not statistically significant. However, the AI-generated explanation of chemotherapy side effects scored higher in scientific accuracy (4.5 vs 3.8, *P*=.004) and satisfaction (4.4 vs 3.7, *P*=.005) compared to non-AI responses..

CONCLUSIONS

This study highlights ChatGPT's effectiveness in providing chemotherapy education, showing that its responses are comparable to traditional sources in scientific accuracy

CONTACT INFORMATION

Moazzam Shahzad MD

moazzamshazad1@gmail.com