

Reliability of ChatGPT in supporting educational microbiology: A study on bacterial species-specific responses

Dinasha WDM¹, Nissanka MTA¹, Fernando WPPD¹, Gunatilake HGD¹, Perera V²

Introduction and Objectives: Chat-Generative Pre-Trained Transformer (ChatGPT), is increasingly used as a conversational tool in microbiological education. It can be used to learn information on microbial species; however, the risk of generating inaccurate content (“hallucinations”) raises concerns about its reliability. This study aimed to evaluate the accuracy of ChatGPT’s responses to general questions on bacterial species, using prompts designed at three language proficiency levels to simulate varying user competencies. Additionally, the study assessed whether the responses included key microbiological terms typically expected in academic or examination settings, where the presence of such keywords is often used as a measure of understanding.

Methods: The bacterial list and questions were finalised and validated by a clinical microbiologist. To mimic a random person asking questions at different language proficiencies, three levels including low, moderate, and high were selected. Each level was represented by three prompts, which were then input into both ChatGPT-3.5 and 4.1-mini models. The results were analysed qualitatively and quantitatively using the eighteenth edition of the Medical Microbiology Guide by David Greenwood and Microbiology Expertise. The results were scored on a scale of accurate, mixed, incomplete, or inaccurate.

Results: There were mean values of 15.57%, 83.87%, 0.53% in 3.5 and 19.58%, 80.21%, 0.19% in 4.1 mini models for accurate, mixed/incomplete and inaccurate answers respectively. According to the mean percentages, the high proficiency level group had a higher percentage of accurate responses than the low and medium proficiency levels. This indicates that, if the question is appropriately stated, the probability of obtaining accurate ChatGPT responses is higher.

Conclusion: ChatGPT provides more accurate answers when questions are clearly and precisely phrased. Inaccurate or vague prompts often lead to partially correct responses. The study highlights the need for clearer questioning and continued model development to improve academic reliability. ChatGPT 4.1 mini is noticeably more accurate than ChatGPT 3.5, offering clearer and more reliable answers across a variety of questions. Further research is needed to validate these findings and compare performance with newer models.

Keywords: ChatGPT, bacterial species, artificial intelligence, microbiology.

¹School of Biomedical Sciences, International Institute of Health Sciences, Welisara, Sri Lanka

²Department of Microbiology, Faculty of Medicine, Sabaragamuwa University of Sri Lanka

Address for correspondence: Mr. Gayan Gunatilake: School of Biomedical Sciences, International Institute of Health Sciences, Welisara, Sri Lanka; Telephone: +94766121311; Email: danushka@iihsciences.edu.lk

 <https://orcid.org/0009-0002-8866-3900>