CME LIVE OPEN ACCESS

CME Journal of Clinical Case Report

Case Report ISSN: 3067-7998

Assessing the Suitability of ChatGpt and Deepseekai For Parents Education on Common Paediatric Respiratory Diseases

Anna Jijo[1], Maria Jijo[2], Swetha Manoj[3*] and Anitha Veerachamy[4]

[1]*Department of paediatrics, Rajiv Gandhi University, Bangalore, Karnataka, India.*

[2]*Department of paediatrics, Apollo Institute of Medical Sciences and Research, Rajiv Gandhi University Hyderabad, Telangana, India.*

[3]*Department of paediatrics, Government Medical College, Kannur University, Kerala, India.*

[4]*Department of paediatrics, Rajiv Gandhi University, Madurai, Tamil Nadu, India.*

Corresponding Author: Swetha Manoj, Department of paediatrics, Government Medical College, Kannur University, Kerala, India.

Received: 2025 July 01 **Accepted:** 2025 July 22 **Published:** 2025 Aug 07

Abstract

Aim

This study aims to evaluate and compare the quality of educational content generated by two artificial intelligence tools—ChatGPT and DeepSeekAI—for educating parents about paediatric respiratory illnesses- croup, bronchiolitis, pneumonia, pertussis.

Methodology

A cross-sectional analysis was conducted using AI-generated educational content. Each response was assessed for word and sentence count, average words per sentence, syllables per word, readability (Flesch Reading Ease Score and Grade Level), similarity percentage (Quillbot) and reliability (modified Discern Score). Statistical analysis was performed using independent sample t-tests, with $p<0.05$ considered significant.

Results

DeepSeekAI responses were longer (mean word count: 422 vs. 333.75) and included more sentences. But no statistically significant differences were found in any variable, including readability (Grade Level: 9.22 vs. 9.30; Ease Score: 41.17 vs. 40.97) and reliability (2.25 vs. 2.00). Similarity scores were also comparable between the two tools (36.02 vs 32.1)

Conclusion

With the increasing use of artificial intelligence (AI) in health communication, tools like ChatgPT and DeepSeekAI offer potential for generating accessible and reliable educational material. Both ChatGPT and DeepSeekAI generated parent education materials of similar quality, readability, and reliability. The findings of this study suggest that AI tools can be utilized for developing parental educational content for common paediatric respiratory conditions.

Key Message

We conducted this study to assess and compare the quality of educational content created by ChatgPT and DeepSeekAI for parent education on common paediatric respiratory illnesses; croup, bronchiolitis, pneumonia, pertussis. Both ChatGPT and DeepSeekAI generated parent education materials of similar quality, readability, and reliability which suggests that they can be used for parental educational content generation.

Keywords: Artificial Intelligence, Chatgpt, Deepseekai, Paediatric Respiratory Diseases, Education

1. Introduction

Paediatric respiratory illnesses such as croup, bronchiolitis, pneumonia, and pertussis are major causes of morbidity and healthcare utilization in children under five years of age. Croup, often caused by parainfluenza viruses, presents with stridor and a characteristic barking cough, frequently leading to emergency visits during fall and winter months. Bronchiolitis, typically associated with respiratory syncytial virus (RSV), is the leading cause of hospitalization in infants under 12 months of age, with peak incidence during the winter season. Pneumonia, whether viral or bacterial, remains a significant cause of respiratory distress and hospitalization, especially in children with underlying comorbidities. Pertussis, caused by Bordetella pertussis, continues to circulate even in vaccinated populations and poses a high risk for severe complications in infants. Parental awareness and early identification of symptoms in these conditions are crucial for timely intervention, effective home care, and reduced hospital admissions. However, low health literacy remains a barrier to effective symptom management and preventive care, particularly in underserved populations [1-4].

With recent advances in large language models (LLMs), artificial intelligence (AI) has emerged as a promising tool to support public health education. ChatGPT-AI (developed by OpenAI) and DeepSeek-AI are two prominent conversational AI platforms that offer personalized, real-time responses to user queries. These tools can aid caregivers in understanding common paediatric symptoms, home care measures, and red flag signs that necessitate medical attention. AI-based platforms are scalable and accessible, making them especially beneficial in low-resource settings. However, limitations such as variable accuracy, lack of context-specific guidance, and potential for misinformation highlight the need for ongoing quality checks and integration with evidence-based guidelines [5,6].

AI-driven patient education has specific utility in the context of paediatric respiratory conditions. For croup, AI platforms can help caregivers distinguish between mild cases managed at home and severe presentations requiring emergency care, especially through the recognition of signs like stridor at rest. In bronchiolitis, ChatGPT-AI and DeepSeek-AI can support parental decision-making regarding hydration, symptom monitoring, and when to escalate care. For pneumonia, these tools may help identify symptoms suggestive of bacterial infection, such as high fever and respiratory distress, guiding families toward timely medical evaluation. In the case of pertussis, AI platforms can reinforce the importance of immunization, early detection of paroxysmal coughing episodes, and the urgency of seeking care for infants at risk of apnea or severe complications. When appropriately designed and validated, AI technologies can complement traditional health systems by enhancing parental counselling, supporting preventive care, and improving child health outcomes.

1.1 Aims and Objectives

This study aims to compare the readability, similarity and reliability of parent education materials generated by ChatGPT and DeepSeekAI for four common paediatric respiratory conditions: croup, bronchiolitis, pertussis, and pneumonia. It aims to draw a comparison based on readability and ease of understanding.

2. Methodology

A cross-sectional original research study was conducted over one week, from March 1 to March 7, 2025. Since no human participants were involved, the study did not require ethical clearance from an institutional review board. The primary objective was to evaluate the readability, originality, and reliability of AI-generated patient education materials on common paediatric respiratory conditions. For this purpose, two advanced AI models were selected: ChatGPT (GPT-4, OpenAI) and DeepSeekAI (DeepSeek-V2). Both tools were tasked with generating educational brochures based on standardized prompts for four paediatric conditions: croup, bronchiolitis, pertussis, and pneumonia. The prompts used were: "Write a patient education guide on [condition]". All responses were generated and collected in Microsoft Word format between March 1 and March 7, 2025. To assess the quality of the generated texts, a multi-step evaluation was conducted. First, readability was evaluated using the Flesch-Kincaid Calculator, which analysed total word count, sentence count, and readability score based on ease of understanding. Second, originality and similarity were assessed using the QuillBot Plagiarism Checker, which scanned each document for overlap with existing content. Third, reliability and quality were graded using a modified DISCERN tool, which is a validated instrument developed to judge the quality of written consumer health information. The DISCERN tool includes five core questions: Are the aims clear and achieved? Are reliable sources of information used? Is the information presented both balanced and unbiased? Are additional sources of information listed for patient reference? Are areas of uncertainty mentioned? Each response was independently rated using this scale to ensure content clarity, trustworthiness, and patient-centeredness [7].

All compiled data were exported to Microsoft Excel and analysed using R version 4.3.2 (R Core Team, 2023). Descriptive statistics were used to summarize readability and reliability scores. Comparisons between ChatGPT and DeepSeekAI outputs were made using an unpaired t-test, with a p-value <0.05 considered statistically significant. Additionally, the relationship between readability and reliability was evaluated using Pearson's Coefficient of Correlation. The comprehensive analysis allowed for an objective comparison of AI-generated health education tools in a paediatric context.

3. Results

ChatGPT and DeepSeekAI were used to generate brochures on patient education of Croup, Figure 1 displays a comparison between ChatGPT and DeepSeekAI on four patient education topics across four key features: grade level, ease score, similarity percent, and reliability score. Table 1 shows the characteristics of responses generated by ChatGPT

and DeepSeekAI. There is no significant difference in the word count (p= 0.309), sentence count (p= 0.365), average words per sentence (p= 0.823), average syllables per word (p= 1.000), grade level (p= 0.940), ease score (p= 0.979), Similarity % (p= 0.643) and reliability score (p= 0.391). In terms of word count, ChatGPT produced a mean of 333.75 (±133.15), while DeepSeek AI produced a slightly higher mean of 422.00 words (±86.37). However, the p-value of 0.309 suggests that the values are not statistically significant and the responses generated are of comparable length overall. In terms of sentence count, ChatGPT produced a mean of 48.50 sentences (±22.84), while DeepSeek AI produced a mean of 62.50 sentences (±17.14). The p value of 0.365 indicates that the values are not statistically significant and that the AI tools are able to present information in a similarly segmented way. In terms of average words per sentence, Chat GPT produced a mean of 7.12 and similarly, Deep Seek AI produced a mean of 6.92. With only a slight difference between the two, a p value of 0.823 indicates the values to not be of statistical significance. Both the models have favoured similar sentence construction. For the average syllables per word which indicates the complexity of the vocabulary used, ChatGPT and Deep Seek AI both produced a mean of 1.87 syllables per word. The p value of 1.000 indicates there is absolutely no difference in this parameter. For the grade level, a mean of 9.30 for ChatGPT and 9.22 for DeepSeekAI was calculated. Based on the p value of 0.940, this indicates that the values are not of any statistical significance. Both brochures are written at the 9th grade level and would be moderately difficult for those with lower health literacy.

The ease score calculated from the responses were also identical: 40.97 for ChatGPT and 41.17 for DeepSeekAI. The scoring system of 0 (very difficult) to 100 (very easy) indicates that the similar scores of 40 produced by the AI tools suggests the texts are fairly difficult to read. A p value of 0.979 suggests that both the models produce similarly readable material. About similarity percentage, ChatGPT produced a mean of 32.10% (±10.51) while DeepSeekAI scored slightly higher at 36.02% (±12.19). The p value of 0.643 indicates that both AI tools produced brochures of similar originality. The reliability score was 2.00 for ChatGPT and 2.25 for DeepSeekAI. A p value of 0.391 indicates no statistical significance and that both the models produce content perceived to be similarly reliable. Overall, both ChatGPT and DeepSeekAI generated brochures with comparable structure, complexity, readability and perceived trustworthiness. Grade level and Ease score were evaluated using the 'Flesch-Kincaid calculator'. Similarity percentage was evaluated by the 'QuillBot Plagiarism Checker and Similarity Tool' and the Reliability score graded through the 'modified DISCERN tool'. In terms of grade level, for croup, ChatGPT produced a grade level of 7.9 and DeepSeek AI a level of 8.4 indicating slightly simpler text in the response produced by ChatGPT. For Bronchiolitis, ChatGPT produced a grade score of 11.9 and DeepSeek AI a score of 9.9 indicating that the response produced by ChatGPT may be harder to read for the average reader. In Pertussis, the values were 8.5 and 9.6 for ChatGPT and DeepSeek AI respectively showing minimal difference. Similar scores of 8.9 for ChatGPT and 9.0 for DeepSeekAI were seen for Pneumonia indicating equal readability. The grade levels produced across all 4 topics indicate material written at the high school level which may be above the average health literacy for some patients. With ease score, ChatGPT consistently scores higher across all four topics. For Croup (ChatGPT: 54, DeepSeekAI: 47.3) and Bronchiolitis (ChatGPT: 42, DeepSeekAI: 22.2), ChatGPT scores significantly higher. For Pertussis (47.1 vs. 38.8), ChatGPT again led slightly while for Pneumonia (40.6 for ChatGPT vs. 40.4 for DeepSeekAI), both values were similar. These values demonstrate that overall, ChatGPT produced more readable content. The similarity percentage reflects originality. A lower similarity score suggests output that is unique. In this regard, ChatGPT consistently produced lower similarity scores across all four topics. For Croup, ChatGPT 's content was more original with 21% compared to a notably high 54% for Deep Seek AI. For Bronchiolitis (ChatGPT: 26.9%, DeepSeekAI: 42%), Pertussis (ChatGPT: 31.9%, DeepSeekAI: 40.1%) and Pneumonia (ChatGPT: 25.3%, DeepSeekAI: 31.3%), similar trends were seen. This suggests that ChatGPT produces more original content. In terms of reliability score, both AI models scored equally across most conditions (on a 3-point scale). For Croup, Deep Seek AI had a slightly higher score of 3 compared to ChatGPT with a score of 2. For Bronchiolitis, Pertussis and Pneumonia, both models were rated equally at a score of reliability of 2. This shows that both AI models produced medically acceptable information based on the reliability score which reflects how trustworthy the content is perceived to be. Although some topic-specific differences appear visually, none of these differences were statistically significant based on t-tests (all $p > 0.05$). This suggests that on average, both AI models produce brochures with similar readability, similarity, and reliability characteristics.

4. Discussion

A cross-sectional study conducted to compare responses generated by the AI tools ChatGPT and DeepSeekAI for brochures on patient education for common Paediatric respiratory conditions such as croup, bronchiolitis, pertussis and pneumonia, revealed that there is no statistically significant difference in the grade level, ease score, similarity percent and reliability score between the two AI tools. Content produced by AI for patient education provides an opportunity for those in resource-limited areas to gain access to health information. It allows patients to make informed and timely decisions on their health. To ensure that information provided in the materials is easy to read and retain by the general public, care should be taken to to keep them simplified and concise. In this study, there is no statistically significant difference between the average words, sentences, words per sentence and syllables per sentence between the two AI tools. The readability of easy-to-read health brochures is recommended as at or below sixth-to-eighth grade level. In this study, the average ease score of the content produced by ChatGPT was 40.97, and that of DeepSeekAI was 41.17, which indicates college grade level. Several studies have been done to assess the role of AI in patient education on health conditions, including a study to understand AI's role in endometriosis patient education.

Citation: Jijo, A., Jijo, M., Manoj, S., Veerachamy, A. (2025). Assessing the Suitability of ChatGpt and Deepseekai For Parents Education on Common Paediatric Respiratory Diseases. *CME J Clin Case Rep, 2*(2), 1-5.

The findings of the study indicated that the responses were accurate with a varying degree of sufficiency [8-11].

AI tools like ChatGPT are trained using massive text datasets in multiple languages, which is then used to provide adequate responses to the input that is given. This can lead to plagiarism. It involves wrongfully presenting someone else's work or ideas as one's own, without having given due credit to the source. Plagiarism is detrimental to medical science as it prevents researchers and medical writers from receiving recognition for their work. In this study, the average similarity percent using ChatGPT was 32.10 and using DeepSeekAI was 36.02. There is no statistically significant difference between the two. A study done on the problem of plagiarism for medical writers highlights the different types of plagiarism as well as potential repercussions in medical research. It explores how plagiarism can be unintentional on the part of the writer, and how to avoid plagiarism in writing [12-14].

The DISCERN tool is an instrument that patients can use to assess the reliability of health information to make their treatment choices. The average DISCERN score for the content produced by ChatGPT was 2, and that of DeepSeekAI was 2.25. There is no statistically significant difference between the two. A study comparing ChatGPT and ChatSonic regarding patient queries on hypertension, indicated variations in the responses, highlighting the need to monitor content produced by AI tools [15,16].

4.1 Limitations

This study is limited by the evaluation of only two AI tools and four diseases. Future research must include a broader range of AI tools and medical conditions to assess the suitability of AI for patient education. Additionally, the version of ChatGPT used in this study is not the most recent, and it cannot be assumed that it provides the latest medical information. Given the rapid advancements in medicine, AI tools may not reflect updated guidelines and practices.

5. Conclusion

This study highlights that there is no statistically significant difference in the average grade level, ease score, similarity percent and and reliability score for the parent education guide generated by ChatGPT and DeepSeekAI on croup, bronchiolitis, pertussis and pneumonia in the paediatric population. Further research must be undertaken to include other AI tools and a variety of medical conditions, including conditions of current public health relevance. It is important to assess whether AI tools used in these studies are provided with up-to-date information and generate content reflecting the same. These tools must be developed to ensure that the output produced is in conjunction with the latest guidelines and can ultimately be verified by the user. Ultimately, AI-generated medical content must be available and accessible to the public at large.

Conflict of Interest

Authors declare no conflict of interest.

References

1. Hall, C. B., Weinberg, G. A., Blumkin, A. K., Edwards, K. M., Staat, M. A., Schultz, A. F., ... & Iwane, M. K. (2013). Respiratory syncytial virus–associated hospitalizations among children less than 24 months of age. *Pediatrics, 132*(2), e341-e348.
2. Jain, S., Williams, D. J., Arnold, S. R., Ampofo, K., Bramley, A. M., Reed, C., ... & Finelli, L. (2015). Community-acquired pneumonia requiring hospitalization among US children. *New England Journal of Medicine, 372*(9), 835-845.
3. Winter, K., Zipprich, J., Harriman, K., Murray, E. L., Gornbein, J., Hammer, S. J., ... & Cherry, J. D. (2015). Risk factors associated with infant deaths from pertussis: a case-control study. *Clinical infectious diseases, 61*(7), 1099-1106.
4. DeWalt, D. A., Broucksou, K. A., Hawk, V., Brach, C., Hink, A., Rudd, R., & Callahan, L. (2011). Developing and testing the health literacy universal precautions toolkit. *Nursing outlook, 59*(2), 85-94.
5. Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. *Journal of medical Internet research, 21*(3), e12802.
6. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine, 25*(1), 44-56.
7. Charnock, D., Shepperd, S., Needham, G., & Gann, R. (1999). DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology & Community Health, 53*(2), 105-111.
8. Ihsan, M. Z., Apriatama, D., & Amalia, R. (2025). AI-assisted patient education: Challenges and solutions in pediatric kidney transplantation. *Patient Education and Counseling, 131*, 108575.
9. Bothun, L. S., Feeder, S. E., & Poland, G. A. (2022). Readability of COVID-19 vaccine information for the general public. *Vaccine, 40*(25), 3466-3469.
10. Kirchner, G. J., Kim, R. Y., Weddle, J. B., & Bible, J. E. (2023). Can artificial intelligence improve the readability of patient education materials?. *Clinical Orthopaedics and Related Research®, 481*(11), 2260-2267.
11. Oliveira, J. A., Eskandar, K., Kar, E., de Oliveira, F. R., & Filho, A. L. D. S. (2024). Understanding Ai's role in endometriosis patient education and evaluating its information and accuracy: systematic review. *Jmir Ai, 3,* e64593.
12. Sallam, M. (2023, March). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare* (Vol. 11, No. 6, p. 887). MDPI.
13. Das, N., & Panjabi, M. (2011). Plagiarism: Why is it such a big issue for medical writers?. *Perspectives in clinical research, 2*(2), 67-71.
14. Helgesson, G., & Eriksson, S. (2015). Plagiarism in research. *Medicine, Health Care and Philosophy, 18*(1), 91-101.

15. Montoya, A., Llopis, N., & Gilaberte, I. (2011). Validation of the translation of an instrument to measure reliability of written information on treatment choices: a study on attention deficit/hyperactivity disorder (ADHD). *Education for Health, 24*(3), 577.

16. Vinufrancis, A., Al Hussein, H., Patel, H. V., Nizami, A., Singh, A., Nunez, B., ... & VINUFRANCIS, A. (2024). Assessing the quality and reliability of AI-generated responses to common hypertension queries. *Cureus, 16*(8).