ORIGINAL ARTICLE

Received: 22/01/2024

Revised: 26/05/2024

Approved: 28/06/2024

Analysis of ChatGPT responses to breastfeeding questions

Natan Viana Medeiros[1], Vitor Fernandes Alvim[2], Maria Teresa dos Santos Silva[2], Sabrine Teixeira Ferraz Grunewald[3]

[1]Universidade Federal de Juiz de Fora. Juiz de Fora (UFJF) – Juiz de Fora (MG), Brazil

[2]Hospital Universitário, Universidade Federal de Juiz de Fora (UFJF) – Juiz de Fora (MG), Brazil

[3]Departamento Materno-Infantil, Faculdade de Medicina, Universidade Federal de Juiz de Fora (UFJF) – Juiz de Fora (MG), Brazil

Corresponding author: Sabrine Teixeira Ferraz Grunewald - Universidade Federal de Juiz de Fora - Avenida Eugênio do Nascimento, s/n° - Dom Bosco – CEP: 36038-330 - Juiz de Fora (MG), Brazil - E-mail: sabrine.pediatria@gmail.com

Declaration of interests: Nothing to declare

Funding: CNPq (401206/2020-3; 303671/2019-0; 314366/2020-2)

https://doi.org/10.7322/abcshs.2023351.2677

ABSTRACT

Introduction: Artificial intelligence tools are impacting medicine in a way that makes knowledge more accessible to both doctors and patients, but their absolute accuracy is still little studied. **Objective:** This study aims to evaluate the quality of responses provided by ChatGPT to potential inquiries from families regarding breastfeeding. **Methods:** A cross-sectional study was conducted through an online questionnaire with active Brazilian pediatricians (n=56) who expressed their opinions on ten pairs of questions and answers related to breastfeeding. Questions were formulated based on common doubts, and responses were obtained after submitting the queries to ChatGPT. The quality of responses was assessed on a scale of 1 to 5 points. **Results:** The findings revealed an average score exceeding 4.0 for all questions posed to the artificial intelligence regarding "Clarity of the provided answer" and "Conformity with current scientific knowledge." Regarding "I am satisfied with the presented response," participants rated ChatGPT responses above 4.0 for most questions. Most pediatricians agreed with the statement, "If I were responding to this question for a real patient, my answer would be different." **Conclusion:** Responses generated by ChatGPT received high satisfaction rates from the population of pediatricians who evaluated them, being considered clear and based on updated scientific knowledge. However, most pediatricians stated they would provide different responses to their patients.

Keywords: Artificial Intelligence; breast feeding; health education.

INTRODUÇÃO

Artificial Intelligence (AI) is a field of study in science that encompasses several areas and seeks to intertwine computer knowledge with human behavior and knowledge[1]. Among AI studies, we have Generative Pre-trained Transformers (GPT), which are systems that can understand and generate long sequences of complex concepts through unsupervised learning by a human[2]. OpenAI is an AI research and implementation company that has created a natural language processing model called ChatGPT. ChatGPT uses the GPT system to respond and produce person-like writing in real time during a conversation that can contain questions, answers, or even venting[2].

The evolution of technology and the internet has impacted medicine in a way that makes knowledge more accessible to both doctors and patients[2]. It is not new to use tools like Google to try to find out symptoms or advice about a disease before seeing a doctor[3]. Patients need to play a leading role in their health process, but the obstacle lies in finding trustworthy places to get information and knowing how to analyze the information they find[4]. With the advent of ChatGPT, we have a new tool that can be used for medical information[5].

According to the literature, parents all over the world are heavy online users of information related to their children's health, in very different circumstances[6]. In studies that evaluated the search for health information online by parents in general, the prevalence of this practice ranged from 52% to 98%, and although parents reported wanting more guidance from their pediatrician on how to find reliable information, they rarely discussed the information they acquired from the Internet with these professionals[6].

Brazil is considered a benchmark for breastfeeding, as since the 1980s, various actions to promote, protect, and support breastfeeding with the participation of the government, society, and non-governmental organizations have contributed to increasing the median duration of breastfeeding from three months in the 1970s to more than one year[7]. Despite progress, the

prevalence of exclusive breastfeeding up to 6 months falls short of the WHO target of at least 70% by 2030[8]. Therefore, it is important to strengthen existing programs, the commitment of health professionals, and the intensification of counseling and health education for breastfeeding women. AI tools such as ChatGPT, despite their potential for health education in various contexts, have not yet been explored in the specific context of breastfeeding.

There are ongoing studies evaluating the reliability of the ChatGPT tool in its responses to patients. An American cross-sectional study that evaluated the quality of four variations of chatbot responses to 26 questions about skin, lung, breast, colorectal and prostate cancer treatment, based on the analysis of four oncologists, resulted in the oncology team recommending that doctors inform patients that AI chatbots are not a reliable source of information about cancer treatment[4]. On the other hand, the cross-sectional study by Ayers and colleagues, also conducted in the USA, showed that patients preferred the answers generated by the chatbot to the doctors' answers to 195 questions on various health topics obtained from online forums, by a ratio of 4 to 1[1]. In addition, the chatbot's responses were rated significantly better in terms of quality and empathy, even when compared to the longer responses from medical authorship[1].

In line with an American study, ChatGPT's answers on cirrhosis and hepatocellular carcinoma were evaluated: the 73 questions were obtained from publications by institutions and societies in the field, as well as social networks supporting patients, and their scientific accuracy was assessed by two experts in the field[9]. As a result, AI provided mostly adequate answers to questions related to basic knowledge, lifestyle habits, and disease treatment but had limitations for reference values, cut-off points, and specific guidelines[9]. Finally, another cross-sectional study in the USA evaluated the tool's responses to 13 myths about cancer, comparing the quality of ChatGPT's responses with those available on the National Cancer Institute's website. The

answers were compared by five reviewers, and 11 of the 13 answers provided by ChatGPT had satisfactory scientific accuracy[10].

Thus, there is still controversy in the literature about the real usefulness and accuracy of this tool for patient health education. In addition, although research is available on the reliability of ChatGPT on a variety of medical subjects, there is still a lack of studies evaluating the effectiveness of this tool in pediatrics and, more specifically, on breastfeeding. Also noteworthy is the lack of Brazilian studies on AI tools and their role in health education, which is relevant because the performance of chatbots can vary according to language.

Thus, this study aimed to evaluate the quality of the answers provided by ChatGPT to potential questions from families about breastfeeding.

METHODS

The research was conducted at the Federal University of Juiz de Fora, but pediatricians from all over Brazil were allowed to participate. Data was collected in August and September 2023.

Pediatricians were invited to take part in the study in a convenience sample. The sample included pediatricians who had completed a medical residency in pediatrics or equivalent specialization, with active registration with the Regional Council of Medicine, who agreed to take part in the study and signed an informed consent form.

Participants who did not complete the data collection tool in its entirety within the time limit of the survey were excluded.

The sociodemographic data collected was age, gender, professional status, as well as information on length of time in the profession and areas of activity.

The questions were designed to represent frequently asked questions from parents about breastfeeding. The questions were written conversationally, for example: “Could my baby be

getting hungry because my milk is weak?" These questions were taken and adapted from a booklet of myths and truths about breastfeeding[11]. They were entered into the ChatGPT 3.5 survey mechanism on August 3, 2023, and the answers obtained were recorded and are available in the Supplementary Material.

The ChatGPT question and answer groups were organized into an online questionnaire, so that each answer could be analyzed by the participating pediatricians according to the following criteria: whether the answer provided is clear to understand; whether the answer provided is in line with current scientific knowledge; whether the professional is satisfied with the answer provided. This assessment was made using a 5-point Likert scale (1 = totally disagree; 5 = totally agree), and the results were presented using the mean and standard deviation of these scores.

In addition, for each ChatGPT question and answer pair, the pediatricians were asked to give their opinion on whether, if they were faced with a real case in which the patient's family had that question, they would provide a different answer to the one provided by ChatGPT. This response was also recorded on a 5-point Likert scale.

The data collected was organized in a Microsoft Excel spreadsheet, and the statistical tests were conducted using the Statistical Package for Social Sciences (SPSS), version 22.0. Descriptive analysis was conducted by calculating frequencies, means, and standard deviations for the variables studied.

This study was approved by the Human Research Ethics Committee of the Federal University of Juiz de Fora, under CAAE number 69939623.4.0000.5147 and Opinion 6,259,662. The participants agreed to the Free and Informed Consent Form.

RESULTS

A total of 56 pediatricians completed the survey, with a predominance of female participants (92.9%), aged between 28 and 72 (average 43.7). Around a third worked only as general pediatricians (33.9%), with 47 (83.9%) pediatricians frequently treating children who are breastfeeding, while 8 (14.3%) occasionally treat them. Table 1 gives details of the sample's sociodemographic data.

In the evaluation of the ChatGPT answers, the score (with possible scores from 1.0 to 5.0) with the pediatricians' opinion was above 4.0 for all the questions presented to the AI for the opinions “The answer provided is clear to understand” and “The answer is in line with current scientific knowledge”. For the option “I am satisfied with the answer provided,” the survey participants also gave a score above 4.0 for most of the ChatGPT answers. Table 2 provides details of these scores.

Finally, for the opinion “If I were answering this question for a real patient, my answer would be different,” for most of the questions, more than half of the pediatricians answered, “I totally agree” or “I partially agree”, as detailed in figure 1.

DISCUSSION

The answers generated by ChatGPT obtained high satisfaction rates from the population of pediatricians who evaluated them. The professionals found the AI answers easy to understand and with an up-to-date scientific basis, and most of them said they would give similar answers to their patients.

This study showed that the AI chatbot generated reliable information in the same way that Pan et al.[3] obtained it. However, this cross-sectional study, carried out in the United States, evaluated the responses of four different chatbots - one of them ChatGPT - to five questions

about cancer, and the answers obtained were judged to be complex language for the general population[3], which was not the case in this study.

Nov et al.[12] selected ten real questions from patients, which were divided up to be answered by ChatGPT or real professionals. In the evaluation of more than four hundred patients, the answers generated by ChatGPT were judged to be like the answers to the same questions answered by healthcare professionals[12]. However, most pediatricians in this study said that they would provide their patients with different answers to those generated by the chatbot. Patients in the study by Ayers et al.[1] reported preferring the answers generated by AI, which were judged to be of higher quality and more empathetic than those provided by doctors.

The pediatricians in the current study judged ChatGPT's answers on breastfeeding to be scientifically dependable. An American cross-sectional survey that assessed the reliability of the tool about eight common colonoscopy questions obtained from hospital websites also found reliable answers in the opinion of four gastroenterologists[5]. On the other hand, other studies available in the literature concluded that the AI chatbot could not be considered a reliable source of scientifically based information on medical doubts and treatments but that it would certainly make a good technical assistant[4,13].

It is important to consider the possibility that ChatGPT performs differently depending on the medical field and the topic of the question. In this study, we evaluated the answers provided on the topic of breastfeeding, for which we have well-established guidelines published by various health societies with uniform guidelines. For example, the guidelines of the Ministry of Health and the Brazilian Society of Pediatrics on breastfeeding are based on the recommendations of the World Health Organization[8]. The tool's behavior may differ in areas that are evolving and changing frequently, such as oncology. One of the reasons for this is that ChatGPT prepares its responses based on information it collects up to a certain date (for the version used in the year 2021), so more recent advances may not be taken into account[10].

Another relevant point is that chatbots consult a huge amount of available information, and the quality of the answers generated will depend on the quality of this information[9]. Also, guidelines on a given subject can often differ between countries, regions, or even institutions, and unless specifically asked to generate answers according to a given guideline, ChatGPT can generate recommendations that differ from what would be recommended for a given clinical scenario[9].

Health education is one of the essential tools for promoting breastfeeding, with guidance for pregnant and breastfeeding women on its importance and the main difficulties with this practice[14]. This targeted follow-up improves adherence to breastfeeding and can be conducted through different strategies[14]. Despite the potential of tools such as ChatGPT to help with patient education, we found no publications in the literature that explored this possibility for topics related to breastfeeding, and thus, the novelty of this research deserves to be highlighted as one of its strengths.

A methodological limitation of this study is the convenience sample of pediatricians, which does not reliably represent the Brazilian population of pediatricians. This is a cross-sectional study, and it is not possible to determine causality between the data obtained. In addition, the chatbot used was the free version of ChatGPT, and there may be changes in information and writing when compared to its paid tool. No information was obtained on the profile of the patients seen by the professionals taking part in the study, nor on whether they work in the public or private sector, which could influence how these professionals assess the suitability of the responses generated by AI for their patients. Finally, it cannot be ruled out that users may ask the same questions in the tool using regionalized language or neologisms, which may influence the answers provided, something that was not addressed in this study.

Many parents have questions about breastfeeding. However, they are unaware of information resources that can provide them with accessible information based on scientific

literature to answer their questions. In addition, the current literature on the implications of AI chatbots in medicine is scarce. Few studies highlight the usefulness of the tool both for healthcare professionals and for patients and users of the system. This study showed that the ChatGPT AI was able to provide information based on current scientific literature about breastfeeding. The tool's answers were accessible and easy to understand for the various doubts presented on the subject.

However, it may still be too early to establish whether ChatGPT, as well as other AI tools, are a reliable source of information for the population, whether on breastfeeding or other topics. It is known that these tools are constantly evolving, adding new features, and correcting flaws. Despite the obvious growth in the potential of these tools and the results of this and other studies showing the accuracy of the information provided, replacing consultation with a health professional still seems a long way off and shouldn't even be an objective. Rather than standing in the shoes of a professional, ChatGPT can complement the information provided, proving to be a useful tool for health education for patients and families.

REFERENCES

1. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589-96. https://doi.org/10.1001/jamainternmed.2023.1838

2. Javaid M, Haleem A, Singh R. ChatGPT for healthcare services: An emerging stage for an innovative perspective. Bench Council Transact Benchmarks Stand Evaluation. 2023;3(1):100105. https://doi.org/10.1016/j.tbench.2023.100105

3. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA Oncol. 2023;9(10):1437-40. https://doi.org/10.1001/jamaoncol.2023.2947

4. Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, et al. Use of artificial intelligence chatbots for cancer treatment information. JAMA Oncol. 2023;9(10):1459-62. https://doi.org/10.1001/jamaoncol.2023.2954

5. Lee TC, Staller K, Botoman V, Pathipati MP, Varma S, Kuo B. ChatGPT answers common patient questions about colonoscopy. Gastroenterology. 2023;165(2):509-11. e7. https://doi.org/10.1053/j.gastro.2023.04.033

6. Kubb C, Foran HM. Online health information seeking parents for their children: systematic review and agenda for further research. J Med Internet Res. 2020;22(8):e19985. https://doi.org/10.2196/19985

7. Estudo Nacional de Alimentação e Nutrição Infantil (ENANI). Aleitamento materno: prevalência e práticas entre crianças brasileiras menores de 2 anos. 4: ENANI - 2019. Rio de Janeiro: UFRJ, 2021.

8. World Health Organization (WHO). United Nations Children's Fund (UNICEF). Global Breastfeeding Scorecard, 2019: Increasing commitment to breastfeeding through funding and improved policies and programmes. Geneva: WHO/UNICEF, 2019.

9. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29(3):721-32. https://doi.org/10.3350/cmh.2023.0089

10. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. JNCI Cancer Spectr. 2023;7(2):pkad015. https://doi.org/10.1093/jncics/pkad015

11. Roque MC. Amamentação: mitos e verdades. Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaíba – Codevasf, 2022, p. 1-32. Available from: https://www.codevasf.gov.br/acesso-ainformacao/institucional/biblioteca-geraldo-rocha/publicações

12. Nov O, Singh N, Mann D. Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. JMIR Med Educ. 2023;9:e46939. https://doi.org/10.2196/46939

13. Garg RK, Urs VL, Agarwal AA, Chaudhary SK, Paliwal V, Kar SK. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. Health Promot Perspect. 2023;13(3):183-91. https://doi.org/10.34172/hpp.2023.22

14. Figueiredo TC, Ribeiro Neto U. Contribuições da educação em saúde na promoção do aleitamento materno. VI Encontro Internacional de Gestão, Desenvolvimento e Inovação. 2022;6(1):1-6.

Table 1: Sociodemographic data of the pediatricians participating in the study (n=56).

	n	**%**
Female	52	92.9
State of operation		
Minas Gerais	40	71.4
Rio de Janeiro	6	10.7
Sergipe	4	7.1
São Paulo	3	5.4
Others	3	5.4
Area of expertise		
General pediatrics	19	33.9
Neonatology	15	26.8
Pediatric Intensive Care	7	12.5
Pediatric Oncology	3	5.4
Other	12	21.4
	Average	Standard Deviation
Age (years)	43.7	11.0
Time since graduating in medicine (years)	18.8	11.4
Time since residency in pediatrics (years)	15.6	11.8

Table 2: Mean and standard deviation of the scores given by pediatricians for the ChatGPT answers to frequently asked questions about breastfeeding.

	Average (standard deviation)		
Doubts submitted to ChatGPT	*The answer provided is clear to understand*	*The answer is in line with current scientific knowledge*	*I am satisfied with the response*
D1. Could my baby be hungry because my milk is weak?	4.8 (0.7)	4.7 (0.6)	4.6 (0.7)
D2. Should the baby be fed every two or three hours when it is only at the breast?	4.5 (0.9)	4.0 (1.3)	3.9 (1.3)
D3. Is it true that breastfeeding makes the breasts fall off?	4.8 (0.4)	4.8 (0.4)	4.7 (0.6)
D4. Do silicone prostheses get in the way of breastfeeding?	4.7 (0.8)	4.6 (0.8)	4.6 (0.9)
D5. Do I need to give my breastfed baby water if it is a sweltering day?	4.9 (0.5)	4.8 (0.5)	4.8 (0.6)
D6. Do I need to wake my baby up at dawn to breastfeed?	4.4 (1.1)	4.3 (1.0)	4.1 (1.2)
D7. Does eating canjica and drinking black beer increase breast milk production?	4.8 (0.4)	4.8 (0.6)	4.7 (0.7)
D8. Until what age can I breastfeed my baby?	4.7 (0.6)	4.5 (0.9)	4.4 (0.9)
D9. Can I drink alcohol while breastfeeding?	4.3 (1.2)	4.0 (1.2)	3.8 (1.4)
D10. If I breastfeed my baby, will it prevent me from getting pregnant?	4.7 (0.7)	4.6 (0.9)	4.5 (0.8)

Figure 1: Frequency of participants' answers to the question "If I were answering this question for a real patient, my answer would be different".