Elevate
Series

ORIGINAL RESEARCH

# Patient Support in Obstructive Sleep Apnoea by a Large Language Model – ChatGPT 4o on Answering Frequently Asked Questions on First Line Positive Airway Pressure and Second Line Hypoglossal Nerve Stimulation Therapy: A Pilot Study

Johannes Pordzik[1], Katharina Bahr-Hamm[1], Tilman Huppertz[1], Haralampos Gouveris [1], Christopher Seifen [1], Andrew Blaikie [2], Christoph Matthias[1], Sebastian Kuhn[3], Jonas Eckrich[1], Christoph R Buhr [1,2]

[1]Department of Otorhinolaryngology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany; [2]School of Medicine, University of St Andrews, St Andrews, UK; [3]Institute for Digital Medicine Philipps-University Marburg and University Hospital of Giessen and Marburg, Marburg, Germany

Correspondence: Johannes Pordzik, Department of Otorhinolaryngology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany, Email johannes.pordzik@unimedizin-mainz.de

**Purpose:** Obstructive sleep apnoea (OSA) is a common disease that benefits from early treatment and patient support in order to prevent secondary illnesses. This study assesses the capability of the large language model (LLM) ChatGPT-4o to offer patient support regarding first line positive airway pressure (PAP) and second line hypoglossal nerve stimulation (HGNS) therapy.
**Methods:** Seventeen questions, each regarding PAP and HGNS therapy, were posed to ChatGPT-4o. Answers were rated by experienced experts in sleep medicine on a 6-point Likert scale in the categories of medical adequacy, conciseness, coherence, and comprehensibility. Completeness of medical information and potential hazard for patients were rated using a binary system.
**Results:** Overall, ChatGPT-4o achieved reasonably high ratings in all categories. In medical adequacy, it performed significantly better on PAP questions (mean 4.9) compared to those on HGNS (mean 4.6) ($p < 0.05$). Scores for coherence, comprehensibility and conciseness showed similar results for both HGNS and PAP answers. Raters confirmed completeness of responses in 45 of 51 ratings (88.24%) for PAP answers and 28 of 51 ratings (54.9%) for HGNS answers. Potential hazards for patients were stated in 2 of 52 ratings (0.04%) for PAP answers and none for HGNS answers.
**Conclusion:** ChatGPT-4o has potential as a valuable patient-oriented support tool in sleep medicine therapy that can enhance subsequent face-to-face consultations with a sleep specialist. However, some substantial flaws regarding second line HGNS therapy are most likely due to recent advances in HGNS therapy and the consequent limited information available in LLM training data.
**Keywords:** AI, ChatGPT, OSA, PAP, hypoglossal nerve stimulation

## Introduction

One-seventh of the world population suffers from obstructive sleep apnoea (OSA) with prevalence predicted to increase in the future.[1] Since OSA is associated with many secondary diseases like arterial hypertension and unfavourable oncologic outcomes for head and neck squamous cell carcinoma, patient support and education is important in order to achieve an early and adequate treatment that is adhered to.[2,3]

Besides the long standing and well-recognised first-line positive airway pressure (PAP) therapy, hypoglossal nerve stimulation therapy (HGNS) has more recently evolved into a promising therapeutical alternative for a specific sub-population of OSA patients. HGNS is now an increasingly used second line therapy for patients not tolerating first line

**2269**

PAP-therapy.[4] Second line therapies are frequently needed in OSA therapy, as PAP therapy shows a non-adherence rate of 34.1%.[5] The most used HGNS system (Fa. Inspire Medical Systems) was approved in 2014 by FDA, showing significant improvements in both objective polysomnography parameters and patient reported outcomes.[6] World-wide, over 75,000 patients have been treated using this second-line approach.[7]

In our sleep medicine consultations, we regularly meet new patients interested in HGNS therapy but who are largely uninformed concerning PAP therapy, HGNS therapy and other second line therapies yet are motivated for more detailed information. However, especially for the newer HGNS therapy, unbiased and evidence-based information is scarce. Crossley et al recently investigated the influence of manufacturers' financial contributions on study results, indicating a lack of independent information.[8] Patients are confronted with advertising from commercial companies on television or in magazines especially for HGNS therapy. The huge amount of unfiltered and possibly biased information regarding different therapies might be overwhelming for patients and can lead to confusion and anxiety regarding how to proceed.[9] Furthermore, the manufacturers' websites offer little information for patients, the common "frequently asked questions" are very limited and devote little or no attention to the side effects of the therapies. (Inspire sleep FAQ 12 questions, Nyxoah FAQ 13 questions and LivaNova no FAQ questions).[10–12] Accordingly, there is a great need for patient information about this new, innovative alternative therapy. Waiting times to see a specialist can be very long. As there is a growing trend of patients turning to artificial intelligence (AI)-driven tools for health-related inquiries, it is therefore reasonable to conclude that patients are seeking information from Large Language Models (LLMs), beyond usual sources of information.

LLMs such as ChatGPT have become a widely used source of information. The introduction of the transformer model by Vaswani et al in 2017 marked a paradigm shift in the field of natural language processing (NLP). While language models previously pursued a word-for-word approach, the transformer model possesses an attention mechanism that allows a high degree of parallelisation. While the base of "knowledge" of previous language models was accumulated by supervised learning, the transformer model is trained by an unsupervised learning approach, thus, representing a true generalist for a wide range of different tasks including healthcare concerns.[13,14] Different studies have shown evidence-based responses to questions commonly asked by patients as well as case-based questions in different medical areas such as orthopaedics,[15] internal medicine,[16] urology[17] and otorhinolaryngology[18,19] among others. However, LLMs still display persistent limitations in information veracity that benefit from the need for experienced specialist judgement.[20] Patients seeking detailed information concerning newer therapeutic options in different fields of medicine will likely "consult" LLMs such as ChatGPT (from OpenAI), Bard (Google), Llama (Meta), and Bing Chat (Microsoft) as a source of information. Especially with regard to the frequently advertised and relatively new second line HGNS patients' demand for information from the internet is increasing. This is illustrated by evidence that shows users searched for "Inspire sleep apnea" on YouTube 2.48 times more between 2018 and 2020.[21]

Whereas the limitations of "googling symptoms on the internet" are largely known, the impact of LLMs on patient education is unknown and subject to a great deal of current research.[22] Therefore, the aim of this study was to investigate whether the LLM ChatGPT-4o, as the most frequently used LLM, is able to correctly answer frequently asked questions about the first line PAP therapy and second line HGNS therapy to support the growing need for patient information in relation to sleep apnoea.

## Materials and Methods

Seventeen questions concerning PAP-therapy were retrieved from literature, and wording was adjusted to the German health system.[23] Same questions were modified to be suitable for the HGNS (see Supplementary Material). Questions were answered using the latest version of ChatGPT-4o (Open AI, San Francisco, United States), respectively. Each answer was rated on a 6-point Likert scale (1 = very poor and 6 = excellent) in the categories medical adequacy, conciseness, coherence, and comprehensibility by experienced experts in sleep medicine.

The key assessment metrics were defined as follow:

- Medical adequacy refers to the accuracy and appropriateness of medical content relative to established clinical guidelines and/or expert consensus.
- Coherence refers to the logical flow and consistency of the information presented. This was assessed by reviewers based on the clarity of connections between ideas and the absence of contradictory statements.

- Comprehensibility measures the ease with which the information can be understood.
- Conciseness evaluates whether the information is sufficiently succinct without omitting critical details.

The evaluation of these categories has proven its effectiveness in previous studies.[17–19] Furthermore, a binary rating on medical completeness of the answers and potential hazard for patients of the information/recommendation provided was assessed. Both the passing of the questions to ChatGPT-4o and the evaluation of the responses took place in June 2024. Statistical analysis was performed using Prism for Windows (version 9.5.1; GraphPad Software) and SPSS 27 (IBM, Armonk, NY, USA).

## Ethical Statement

There is no need for any specific ethical approval in this kind of studies not involving patients, animals or cells. All right, title, and interest, if any, in and to Output are assigned to the customer by ChatGPT-4o (Open AI, San Francisco, United States) (see terms of use: ownership of content: https://openai.com/policies/row-terms-of-use/). All involved sleep experts rating the answers are mentioned authors.

## Results

Figure 1 visualises the rating of the 3 experienced experts in sleep medicine for the categories medical adequacy, coherence, comprehensibility and conciseness. Regarding medical adequacy, ChatGPT-4o answers on PAP questions (mean 4.9; range [1–6]) were significantly rated higher compared to answers on HGNS questions (mean 4.6; range [2–6]) ($p < 0.05$). On coherence, ChatGPT-4o achieved a mean rating of 5.3 (range [1–6]) for PAP answers and 5.3 (range [2–6]) for HGNS answers (ns > 0.05). In the category comprehensibility, ChatGPT-4o performed slightly better on PAP questions with a mean of 5.7 (range [4–6]) compared to HGNS questions with a mean of 5.5 (range [4–6]) (ns > 0.05). In terms of conciseness, ChatGPT-4o was rated mean 5 (range [3–6]) on answers for PAP and mean 4.8 (range [3–6]) on HGNS answers (ns > 0.05). Interrater reliability is shown in Table 1.

Regarding the binary rating for completeness, ChatGPT-4o´s answers on PAP were considered as complete in 45 of 51 ratings (88.24%), whereas 28 of 51 ratings (54.9%) for HGNS answers confirmed completeness of the response. A potential hazard for patients was indicated in 2 of 52 ratings (0.04%) for PAP answers and none for HGNS answers. ChatGPT-4o used a mean amount of 2132 characters [1336–3132] on PAP answers compared to 2020 characters [971–4178] on HGNS questions (ns > 0.05) (Figure 2).

## Discussion

Previous studies assessed the application of ChatGPT for clinical use in Otorhinolaryngology and sleep medicine.[25–30] However, in our outpatient office, we often experience patients not tolerating first line PAP therapy and highly motivated to access unbiased information on HGNS therapy. While confronted with long waiting periods for appointments with sleep specialists, patients are often exposed to advertising by HGNS manufacturers with high potential for bias. Within this study, we evaluated ChatGPT-4o as a tool to provide patients more impartial and balanced information on "frequently asked questions" on HGNS as well as PAP therapy.

Our data show reasonable ratings for ChatGPT-4o´s responses to frequently asked questions about PAP and HGNS therapy in the categories of medical adequacy, conciseness, coherence, and comprehensibility. Although the ratings for PAP answers were rated higher in each category by experienced experts in sleep medicine, only medical adequacy showed a significant difference between PAP (mean 4.9 [1–6]) and HGNS responses (mean 4.6 [2–6]) ($p < 0.5$). ChatGPT-4o achieved similar scores for PAP and HGNS answers in the other 3 categories of conciseness, coherence and comprehensibility, ranging from a mean of 5.3 (both PAP and HGNS) in coherence to a mean of 5.7 (PAP) and 5.5 (HGNS) for comprehensibility. Answers were rated as complete in 45 of 51 ratings (88.24%) for PAP and in 28 of 51 ratings (54.9%) for HGNS.

Studies investigating "frequently asked questions" about specific OSA therapy are uncommon. One study demonstrated the potential of ChatGPT in the management of obstructive sleep apnea but mainly focused on a comparison between sleep surgeons' skills and Chat-GPT.[27] Campbell et al showed that ChatGPT provides appropriate answers to
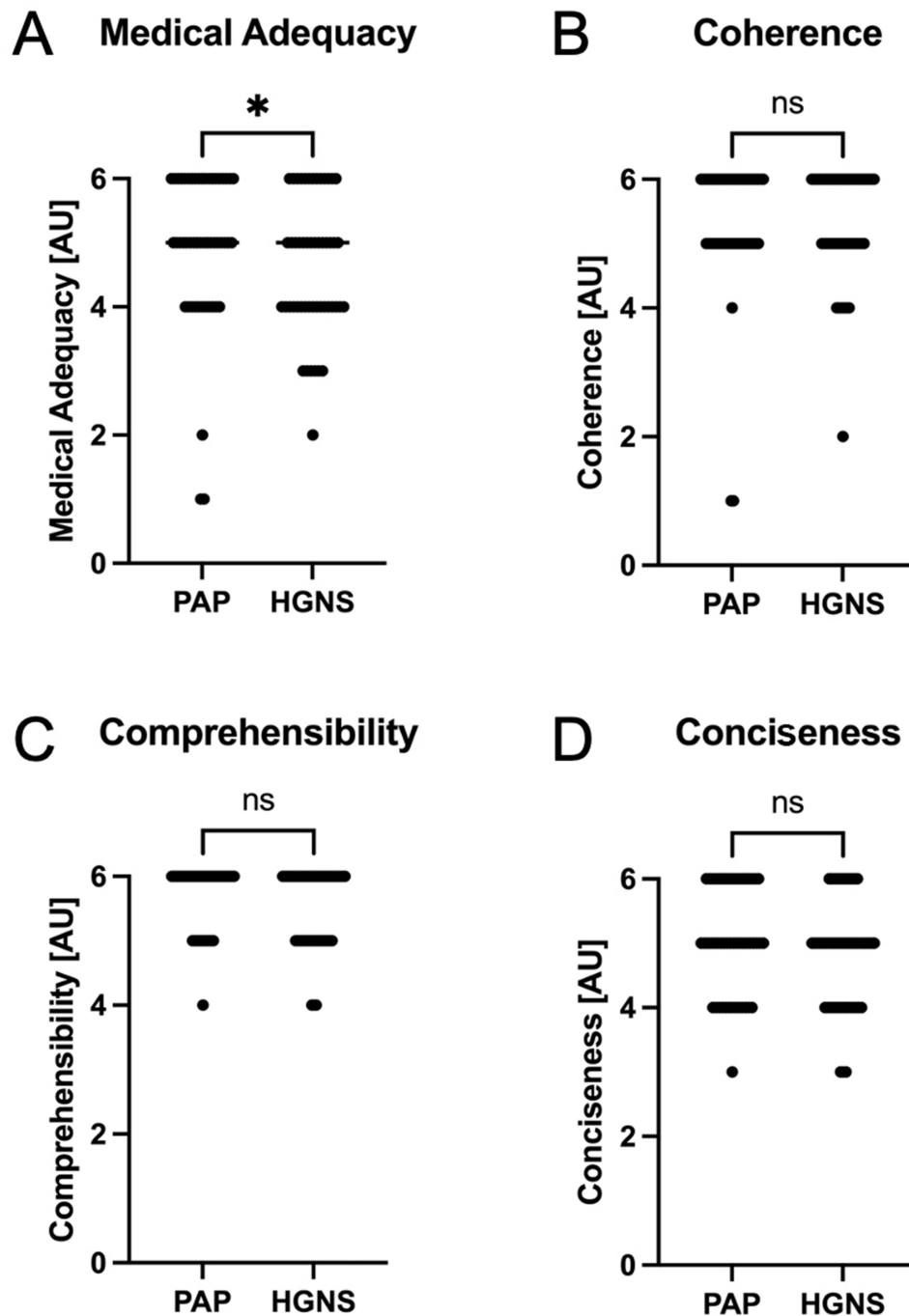
**Figure 1** Comparison of answers given by ChatGPT-4o for questions regarding PAP and HGNS respectively: Rating on a 6 point Likert-scale for (**A**) medical adequacy; (**B**) coherence; (**C**) comprehensibility and (**D**) conciseness. Data shown as a scatter dot blot with points representing absolute values and bar width representing amount of individual values. Horizontal lines represent mean (95% CI). Normality distribution was tested with the D'Agostino and Pearson test. Comparisons were performed using the Wilcoxon test. ns > 0.05, *p < 0.05.

most questions on OSA.[28] A separate study evaluated ChatGPT and Google Bard's ability to inform patients about obstructive sleep apnea (OSA) and its treatment.[29] ChatGPT demonstrated a high level of understanding, correctly answering 90.78% of "frequently asked questions" about OSA treatment. These findings align with our study's comprehensibility ratings, which showed a score of 5.7 (range 4–6) for positive PAP therapy and 5.5 (range 4–6) for HGNS questions. However, it is important to note that the comparative study focused more on general OSA therapy strategies rather than providing detailed information specific to PAP and HGNS therapies.

**Table 1** Interrater Reliability

| | | Fleiss Kappa Question 1–17 (Rater 1–3) | p value |
|---|---|---|---|
| Completeness of the response | PAP | 0.62 | < 0.001 |
| | HGNS | 0.53 | < 0.001 |
| Potential hazard for patients | PAP | 0.48 | < 0.001 |
| | HGNS | 0.53 | < 0.001 |
| | | Kendall W Question 1–17 Rater (1–3) | |
| Coherence | PAP | 0.12 | 0.13 |
| | HGNS | 0.59 | < 0.001 |
| Comprehensibility | PAP | 0.30 | 0.01 |
| | HGNS | 0.05 | 0.40 |
| Conciseness | PAP | 0.61 | < 0.001 |
| | HGNS | 0.68 | < 0.001 |
| Medical adequacy | PAP | 0.26 | 0.01 |
| | HGNS | 0.60 | < 0.001 |

**Note**: Interrater reliability was proved by Fleiss Kappa for nominal skaled data and Kendalls W for ordinal skaled data; interpretation Fleiss Kappa/ Kendall W: <0.00 = Poor; 0.00–0.20 = Slight; 0.21–0.40 = Fair; 0.41–0.60 = Moderate; 0.61–0.80 = Substantial; 0.81–1.00 = Almost perfect; statistical significant p < 0.05[24].

In our study, ChatGPT-4o achieved significantly higher scores on medical adequacy for answers regarding PAP therapy compared to HGNS therapy. This is not surprising, as PAP therapy has been the standard therapy in OSA treatment for many years, with much more information available for training LLMs.[31] On 07/07/2024 PubMed revealed 21,212 results (07/07/2024) for the search query "positive airway pressure therapy" compared to 633 results for "hypoglossal nerve stimulation therapy". Consequently, there are 33.5 times more articles listed in PubMed for PAP compared to HGNS therapy. It was interesting to note that ChatGPT-4o failed to refer to two very recently released HGNS systems (Nyxoah SA and Livanova Inc.) reflecting the nature of how LLMs learn. The limitations in ChatGPT-4o's ability to account for recently released HGNS systems underscore the importance of recognizing the static nature of its training data and the potential for knowledge gaps as new advancements emerge. This highlights the need for continuous monitoring and periodic re-evaluation of LLMs/AI tools to ensure their outputs remain accurate and relevant in dynamic fields such as healthcare.

As language models advance, integrating mechanisms for real-time updates or access to external, up-to-date databases may mitigate these limitations and enhance their reliability. Furthermore, our findings emphasize the necessity for researchers, practitioners and potentially also patients/the general public to critically appraise AI-generated content. Especially for scientists, it is important to validate them against the most current evidence available.

In our study, ChatGPT-4o also demonstrated limitations in understanding MRI-compatibility and testing procedures for HGNS implant patients and showed bias towards specialized medical departments offering HGNS therapy. This suggests ChatGPT-4o has access to more comprehensive information on PAP therapy than HGNS therapy, potentially leading to greater medical accuracy for PAP-related queries. This is consistent with our observation that ChatGPT-4o's answers were more complete for PAP (88.24%) compared to HGNS (54.9%).
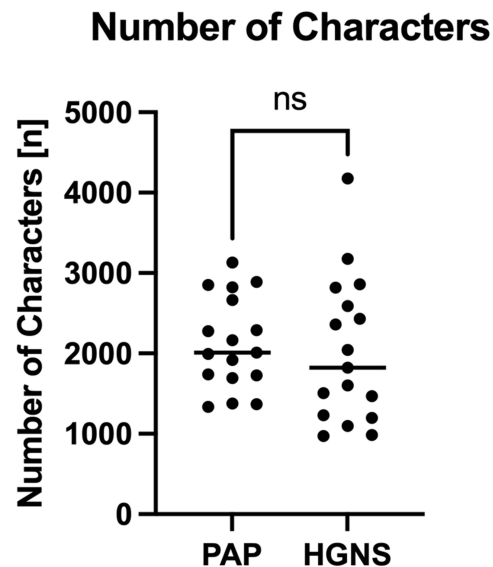
## Number of Characters



**Figure 2** Number of characters used by ChatGTP-4o for answers regarding PAP and HGNS respectively. Data shown as a scatter dot blot with each point resembling an absolute value. Horizontal lines represent the median. Normality distribution was tested with the D'Agostino and Pearson test. Group comparison was performed using the paired *t*-test. ns > 0.05.

The conciseness scores show the powerful ability of ChatGPT-4o to summarise abundant information on both PAP and HGNS therapy. This is further highlighted by the similar length of the responses for PAP therapy and HGNS therapy (ns > 0.05). Only one answer of ChatGPT-4o on PAP therapy was considered as potentially hazardous for patients. In this case, ChatGPT-4o confused the term PAP equipment with the German abbreviation PSA ("Persönliche Schutzausrüstung") which refers to personal protective equipment for working staff. This led ChatGPT-4o to recommend patients to obtain PAP equipment in do-it-yourself stores, specialist occupational safety retailers or online shops. Moreover, ChatGPT-4o states that "in many professions, the employer provides the necessary protective equipment. It is therefore worth asking the employer whether and what equipment is provided". Just because the confusion of ChatGPT-4o is obvious and the answer is wrong, many users may realise this is misleading.

The misunderstanding by ChatGPT-4 likely stemmed from the question being asked in German. While this specific error might not have occurred in English, it highlights the vulnerability to misinterpretations in less common languages. This phenomenon, often called "intelligence hallucination", occurs when language models generate incorrect, unfounded answers instead of admitting their inability to solve a task accurately.[32] This observation underscores a key limitation of current AI chatbots, particularly in fields where context-specific nuances are critical for accurate information. The confusion caused by ambiguous abbreviations like PAP demonstrates the potential for LLMs to misinterpret or over-simplify information. This can lead to misinformation, which is especially concerning in patient education, where clarity and accuracy are paramount. To address these challenges, we emphasize the necessity of extensive testing and validation of AI-generated outputs, particularly in multilingual and domain-specific contexts. Future developments should focus on incorporating mechanisms to recognize contextual ambiguities and prompt the user for clarification when potential errors are detected. Additionally, hybrid models that integrate LLMs with curated, domain-specific knowledge bases (eg grounding the LLM chatbot to certain information sources) could help mitigate such inaccuracies. Rigorous evaluation protocols and ongoing refinement processes should be prerequisites before deploying AI chatbots in sensitive areas like patient education.

However, LLMs are already used by patients and may be considered as official information resources especially for new therapies as patient information about HGNS is lacking. At the same time, LLMs provide a convenient and low barrier resource for accessing comprehensible information in plain language.

There are four main limitations of our study. First, no patients were involved. Even if the assessment of coherence and medical adequacy should be reserved for sleep medicine specialists, patients' assessment of comprehensibility and

conciseness would be of value. So far, one study demonstrated that patients preferred the answers to frequently asked questions about cPAP therapy from pulmonologist experts compared to ChatGPT responses.[30] However, a major problem is that the answers must first be checked for medical accuracy. Patients cannot judge whether the answers to the questions are medically correct. In this respect, our pilot study focuses on whether ChatGPT-4o is able to present medically adequate, coherent, comprehensible and concise answers at all. Future studies should focus on LLMs based answers to FAQs after checking medical accuracy by medical experts. Answers that have not been checked for medical accuracy could confuse patients and lead to serious misunderstandings. Furthermore, due to data protection, studies assessing patients' perspective should evaluate this aspect using local running LLMs without any internet connection ensuring secure local data processing without risking unauthorized data drain. Second, ChatGPT-4o was only asked in German and had to answer in German. English language was used in answering one question by ChatGPT-4o despite the question being asked in German. Answers of ChatGPT-4o to frequently asked questions concerning PAP and HGNS therapy may differ with the language used. However, existing literature highlights ChatGPT's effectiveness as a highly capable translation tool, which helps mitigate potential language-related limitations.[33] Furthermore, most people will likely use LLMs such as ChatGPT-4o in their native language, which in most cases will not be English. In this pilot study, we aim to depict a realistic scenario from our patients' perspective in a non-English-speaking country. Third, a benchmark for FAQ's answers for a sufficient comparison is missing. Unfortunately, the internet provides little unbiased information regarding patient PAP queries and especially HGNS queries. Manufacturers' websites showcase only a few "frequently asked questions" devoting little or no attention to the potential risks and side effects of the therapies. (Inspire sleep FAQ 12 questions, Nyxoah FAQ 13 questions and LivaNova no FAQ questions).[10–12] Moreover, Crossley et al recently investigated the influence of manufacturers' financial contributions on study results, indicating a lack of independent information.[8] Therefore, a reliable benchmark is missing. Our benchmark was the expertise of our medical experts, who rated the answers given by ChatGPT-4o. Fourth, both answers came from an AI and the raters were familiar with the project. This may have biased the rating.

Despite these limitations, this study showcases the great potential of LLMs for patient education on PAP and HGNS therapy. Assuming further improvements, tools such as ChatGPT could be useful to clarify general patient questions in advance of medical consultations. In this case, doctor's consultation could be used more efficiently for difficult and individual patient queries. Moreover, better informed patients may also have the potential to favor shared-decision making. Though, before an official use can be recommended, security shortcomings such as above mentioned hallucinations must be overcome. Nonetheless, doctors should be aware that patients use online resources and LLMs such as ChatGPT before clinical consulting. Doctors should therefore take patients' knowledge carefully into account, as patients' knowledge and resulting expectations could have a relevant impact on the doctor–patient relationship.[34] Accordingly, it is even more important for doctors to be familiar with LLMs as a source of patient education.

## Conclusion

To the best of our knowledge, this is the first study assessing ChatGPT-4o on answering frequently asked questions about first line PAP therapy and second line HGNS therapy. Our study shows the great potential of ChatGPT-4o as an information source for patients with OSA. Consistent with the above mentioned previous studies, ChatGPT-4o demonstrates promising results in educating patients about medical information. Although ChatGPT-4o showed several limitations regarding second line HGNS therapy, in the near future this is likely to improve. Current shortcomings and barriers to introducing LLMs in clinical practice include concerns about data privacy, accuracy and reliability of AI-generated insights and the need for rigorous validation and regulation to ensure patient safety.[30] However, ChatGPT-4o and potentially other LLMs will become highly accessible and powerful support and education tools for lay people, especially for augmenting subsequent consultations with specialists. Future studies should assess the performance of LLMs in real patient interactions on a greater scale.

## Abbreviations

OSA, Obstructive sleep apnoea; LLM, Large language model; PAP, Positive airway pressure; HGNS, Hypoglossal nerve stimulation; NLP, Natural language processing; AI, Artificial Intelligence.

## Data Sharing Statement

The data underlying this article cannot be shared publicly for the privacy of individuals that participated in the study. The data will be shared on reasonable request to the corresponding authors.

## Ethics Approval and Consent to Participate

There is no need for any specific ethical approval in this kind of studies not involving patients, animals or cells.

## Disclosure

SK is the founder and shareholder of MED digital. The authors report no other conflicts of interest in this work.

## References

1. Lyons MM, Bhatt NY, Pack AI, Magalang UJ. Global burden of sleep-disordered breathing and its implications. *Respirology*. 2020;25(7):690–702. doi:10.1111/resp.13838
2. Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. *New Engl J Med*. 2000;342(19):1378–1384. doi:10.1056/NEJM200005113421901
3. Huppertz T, Horstmann V, Scharnow C, et al. OSA in patients with head and neck cancer is associated with cancer size and oncologic outcome. *Europ Archiv Oto-Rhino-Laryngol*. 2021;278(7):2485–2491. doi:10.1007/s00405-020-06355-3
4. Steffen A, Heiser C, Galetke W, et al. Die Stimulation des Nervus hypoglossus in der Behandlung der obstruktiven Schlafapnoe–Aktualisiertes Positionspapier der Arbeitsgemeinschaft Schlafmedizin der DGHNO-KHC. *Laryngo-Rhino-Otologie*. 2021;100(01):15–20. doi:10.1055/a-1327-1343
5. Rotenberg BW, Murariu D, Pang KP. Trends in CPAP adherence over twenty years of data collection: a flattened curve. *Int. J. Otolaryngol*. 2016;45 (1):43. doi:10.1186/s40463-016-0156-0
6. Strollo PJ Jr, Soose RJ, Maurer JT, et al. Upper-airway stimulation for obstructive sleep apnea. *N Engl J Med*. 2014;370(2):139–149. doi:10.1056/ NEJMoa1308659
7. Inspire sleep apnea innovation. Available from: https://www.inspiresleep.de/inspire-therapie/zungenschrittmacher-voraussetzungen-op-ablauf-kosten/?gad_source=1&gclid=CjwKCAiA9IC6BhA3EiwAsbltOA9TFmDRy4N862EO6YaAzLSem4xvJVmbERgs3OowvFo-wS4f7bbcqxoCArMQAvD_BwE. Accessed November 22, 2024.
8. Crossley JR, Wallerius K, Hoa M, Davidson B, Giurintano JP. Association between conflict of interest and published position on hypoglossal nerve stimulation for sleep apnea. *Otolaryngology*. 2021;165(2):375–380. doi:10.1177/0194599820982914
9. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. *J Inf Sci*. 2008;35(2):180–191. doi:10.1177/0165551508095781
10. Inspire Sleep Apnea Innovation/ häufig gestellte Fragen. [frequently asked questions]. Available from: https://www.inspiresleep.de/inspire-therapie /haeufig-gestellte-fragen/. Accessed November 22, 2024.
11. GENIO by Nyxoah/ häufig gestellte Fragen. [frequently asked questions]. Available from: https://www.geniosleep.com/de/faqs. Accessed November 22, 2024.
12. LivaNova health innovation that matters. Available from: https://www.livanova.com/en-us/obstructive-sleep-apnea. Accessed November 22, 2024.
13. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. *OpenAI blog*. 2019;1(8):9.
14. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:1.
15. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. *J Bone Hoint Surg Am*. 2023;105(19):1519–1526.
16. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023;29(3):721–732. doi:10.3350/cmh.2023.0089
17. Eckrich J, Ellinger J, Cox A, et al. Urology consultants versus large language models: potentials and hazards for medical advice in urology. *BJUI Compass*. 2024;5(5):438–444. doi:10.1002/bco2.359
18. Buhr CR, Smith H, Huppertz T, et al. ChatGPT versus consultants: blinded evaluation on answering otorhinolaryngology case–based questions. *JMIR medl edu*. 2023;9(1):e49183. doi:10.2196/49183
19. Buhr CR, Smith H, Huppertz T, et al. Assessing unknown potential—quality and limitations of different large language models in the field of otorhinolaryngology. *Acta Oto-Laryngologica*. 2024;2024:1–6.
20. Daungsupawong H, Wiwanitkit V. ChatGPT versus medical professionals: comment. *Health Serv. Insights*. 2024;17:11786329241241904. doi:10.1177/11786329241241904
21. Xiao K, Campbell D, Mastrolonardo E, Boon M, Huntley C. Evaluating YouTube videos on hypoglossal nerve stimulation as a resource for patients. *Laryngoscope*. 2021;131(11):E2827–E2832. doi:10.1002/lary.29809
22. Tang H, Ng JH. Googling for a diagnosis--use of Google as a diagnostic aid: internet based study. *BMJ*. 2006;333(7579):1143–1145. doi:10.1136/ bmj.39003.640567.AE
23. Chokroverty S. *Questions & Answers About Sleep Apnea*. Jones & Bartlett Publishers; 2009.
24. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Administrative Pharm*. 2013;9(3):330–338. doi:10.1016/j.sapharm.2012.04.004
25. Maksimoski M, Noble AR, Smith DF. Does ChatGPTChat GPT answer otolaryngology questions accurately? *Laryngoscope*. 2024;134 (9):4011–4015. doi:10.1002/lary.31410
26. Buhr CR, Smith H, Huppertz T, et al. Assessing unknown potential-quality and limitations of different large language models in the field of otorhinolaryngology. *Acta Otolaryngol*. 2024;144(3):237–242. doi:10.1080/00016489.2024.2352843

27. Mira FA, Favier V, Dos Santos Sobreira Nunes H, et al. Chat GPT for the management of obstructive sleep apnea: do we have a polar star? *Europ Archiv Oto-Rhino-Laryngol*. 2024;281(4):2087–2093. doi:10.1007/s00405-023-08270-9

28. Campbell DJ, Estephan LE, Mastrolonardo EV, Amin DR, Huntley CT, Boon MS. Evaluating ChatGPT responses on obstructive sleep apnea for patient education. *J Clin Sleep Med*. 2023;19(12):1989–1995. doi:10.5664/jcsm.10728

29. Cheong RCT, Unadkat S, McNeillis V, et al. Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: chatGPT versus Google Bard. *Europ Archiv Oto-Rhino-Laryngol*. 2024;281(2):985–993. doi:10.1007/s00405-023-08319-9

30. Martini A, Ielo S, Andreani M, Siciliano M. ChatGPT: friend or foe of patients with sleep-related breathing disorders? *Sleep Epidemiol*. 2024;4:100076. doi:10.1016/j.sleepe.2024.100076

31. Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. *Lancet*. 2014;383(9918):736–747. doi:10.1016/S0140-6736(13)60734-5

32. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: implications in Scientific Writing. *Cureus*. 2023;15(2):e35179. doi:10.7759/cureus.35179

33. Lee TK. Artificial intelligence and posthumanist translation: chatGPT versus the translator. *Appl. Linguist Rev.* 2023;15:2351–2372.

34. Caiata-Zufferey M, Abraham A, Sommerhalder K, Schulz PJ. Online health information seeking in the context of the medical consultation in Switzerland. *Qualitative Health Research*. 2010;20(8):1050–1061. doi:10.1177/1049732310368404