GENERATIVE AI IN ASSISTING PATIENTS WITH SYRINGOMYELIA AND HYDROCEPHALUS: A PRELIMINARY COMPARISON OF CHATGPT AND GEMINI

PAVLINA KRÖCKEL[1,2]
[1] Friedrich-Alexander Universität Erlangen-Nürnberg, Nuremberg, Germany
pavlina.kroeckel@fau.de
[2] Universitäres Hochschulinstitut Schaffhausen, Schaffhausen, Switzerland
pavlina.kroeckel@fau.de

Patients have been relying on online resources for more information on their symptoms and diagnosis. Existing research has examined Generative AI (GenAI), mostly via ChatGPT as a way of providing more information or even comfort to patients. However, research is scarce on the appropriateness of GenAI to support and inform patients with rare conditions. These patients often find themselves in a difficult-to-navigate maze especially because they have never heard of their condition before the diagnosis and the information online can be overwhelming. In this pre-study, we demonstrate the potential of GenAI to provide support to patients concerning their diagnoses. We work with a patient who has two rare neurological conditions – syringomyelia and hydrocephalus. Utilizing a qualitative and quantitative methodology, including the Patient Education Materials Assessment Tool for Printable Materials (PEMAT-P) and the Flesch Kincaid Reading Ease (FRE) score, we analyzed the patient's feedback to a series of prompts derived from their medical reports. The results demonstrate the potential of generative AI as a valuable support for patients, with preliminary findings indicating that Gemini excels in certain aspects such as understandability, actionability, readability, and is more trustworthy, making it an effective tool for patient education in this context.

1   Introduction

Patients with rare disorders, such as syringomyelia and hydrocephalus, have specialized information needs owing to the complexity of their conditions. These needs include access to high-quality, comprehensible information, technologies that facilitate communication among various stakeholders, and patient portals that centralize resources to enhance usability and empower decision making (Litzkendorf et al., 2016; Long et al., 2022; Stanarević Katavić, 2019). Since its launch in 2022, ChatGPT has significantly influenced AI's role in healthcare, particularly in patient education and medical decision making (Campbell et al., 2023; Chiesa-Estomba et al., 2023; Gabriel et al., 2023; Gordon et al., 2024). However, the application of digital tools such as ChatGPT and Gemini for patients with syringomyelia or hydrocephalus is still not well documented. Our study investigates how these Generative AI tools assist patients in understanding and managing their health conditions amidst the overwhelming amount of online information and limited resources from healthcare providers.

Hydrocephalus is characterized by abnormal circulation of cerebrospinal fluid, leading to symptoms such as headaches and disorientation, with shunting as the primary treatment for draining excess fluid (Bristol, 2014; Hydrocephalus Association, 2024). Syringomyelia involves a fluid-filled cyst within the spinal cord, causing varying symptoms and potential nerve damage, which complicates both diagnosis and treatment (Milhorat, 2000; Zheng et al., 2023). It is rare, but possible for a patient to be diagnosed with both conditions.

Our study focuses on a 38-year-old female initially diagnosed with hydrocephalus at the age of 33 years, required emergency shunt surgery, and later developed syringomyelia due to complications from spinal anesthesia. Unfamiliar with their conditions, they turned to online sources to seek clarity, comfort, and alternative treatments, as well as to understand the complex medical jargon in the reports.

This research aims to answer the question: "*Can Generative AI tools like ChatGPT or Gemini help patients with syringomyelia and hydrocephalus understand their condition and navigate their healthcare journey*?" By addressing this question, we intend to explore the potential of these tools in simplifying complex medical information and aiding treatment decision-making, thus filling a gap in the current research on these conditions.

P. Kröckel: Generative AI in Assisting Patients with Syringomyelia and Hydrocephalus: A Preliminary Comparison of Chatgpt and Gemini

673

2 Methodology

We initiated our study by interviewing the patient to gather essential information about their diagnosis and treatment, forming questions based on prior research and online patient forum discussions. After identifying their main challenge as "understanding their medical reports", we used an MRI report from July 7, 2023, filled with complex medical terminology, to generate specific questions. These questions, listed in Table 1, were categorized into four themes: 1) **_Interpretation_** – this prompt referred to interpreting the results from the radiology report; 2) **_Explanation_** – which included additional information for the chatbot, that is, that the patient has both hydrocephalus and syringomyelia (to test whether this affects the interpretation); 3) **_Treatment_** – we asked for suggestions on how to slow down the worsening of the condition, especially alternative medicine treatments; and 4) **_Reliability & Trustability_** – evaluating the answers in terms of how trustable and reliable they seem. A full list of questions is provided in the supplementary file[1].

Both ChatGPT and Gemini were tested using these prompts to determine their effectiveness in helping the patient navigate their health information. The patient used both AI tools via their account, and detailed results are available in the supplementary file.

Table 1: Terms extracted from the radiology report

Original sentences from the report	Translated in English
Chondrotische Veränderungen mit Protrusionen C3 bis C7	Chondrotic changes with protrusions at C3 to C7
Signalverlust des Nucleus pulposus der Bandscheiben in der T2-Wichtung mit diskreten Bandscheibenvorwölbungen C3 bis C7 zum Teil auch leicht paramedian betont	Loss of signal intensity of the nucleus pulposus of the intervertebral discs in T2-weighted imaging with discrete disc protrusions at C3 to C7, partly also slightly paramedian accentuated
Sklerosen der kleinen Wirbelkörpergelenke	Sclerosis of the small vertebral facet joints
Eintauchen der Kleinhirntonsillen in das Foramen magnum	Displacement of the cerebellar tonsils into the foramen magnum

---

[1] The file can be downloaded here.

We conducted a **qualitative** evaluation by using methods similar to those described by Pate et al. (2021). **Quantitatively**, we assessed the results using the Patient Education Materials Assessment Tool for Printable Materials (PEMAT-P), as referenced in studies by Ayoub et al. (2023), Coskun et al. (2023), McCarthy et al. (2023). For readability, we utilized the WebFX tool to calculate the Flesch Kincaid Reading Ease (FRE) score, a method established in 1975 for the US Navy (Kincaid et al., 1975) and commonly used to measure the readability of medical texts (Bernard et al., 2018; Szmuda et al., 2020).

3   Preliminary Results

Understandability and actionability

The patient used the PEMAT Tool and assigned scores for both criteria. Several areas were evaluated within these criteria. The results are presented in Table 2.

Table 2: PEMAT score comparison for ChatGPT & Gemini

Criteria	Area	ChatGPT	Gemini
Understandability	Content	2	2
	Word choice & Style	1	3
	Organization	4	4
	Layout & Design	1	1
Actionability	Action identification	1	1
	Direct user address	0	1
	Action breakdown	0	1

Note: In the PEMAT tool, it is possible to assign either 0 (disagree) or 1 (agree), or for some areas, an NA (not applicable).

To assess the understandability and actionability of the information provided by ChatGPT and Gemini, several criteria were evaluated. Both tools scored 2/2 on content, indicating that the information was understandable to the patient. However, there was a significant difference in word choice and style; Gemini achieved a perfect score of 3/3, suggesting its use of a simpler, more patient-appropriate language, whereas ChatGPT scored only 1/3, reflecting its use of more complex language that might hinder patient comprehension. Both tools excelled in organization, each scoring 4/4, which signifies that the information was well structured and thus

P. Kröckel: Generative AI in Assisting Patients with Syringomyelia and Hydrocephalus: A Preliminary Comparison of Chatgpt and Gemini

675

facilitated understanding. In addition, the layout and design of both tools were deemed satisfactory, with each scoring 1/1.

Regarding **actionability**, both ChatGPT and Gemini demonstrated effectiveness by scoring 1/1 in action identification, indicating that they provided clear actionable advice for the patient. However, in terms of direct user address, Gemini outperformed ChatGPT with a score of 1/1 compared to ChatGPT's 0, showing Gemini's superior ability to engage directly with the user. Furthermore, Gemini proved more effective in breaking down actions into clear, manageable steps, scoring 1/1, whereas ChatGPT did not score in this area. Gemini's capability enhances both the understandability and execution of suggested actions, making it a more effective tool for patient education in this context.

Readability

As previously mentioned, we used the FRE score to evaluate this criterion. A comparison of the readability scores is shown in Figure 1.
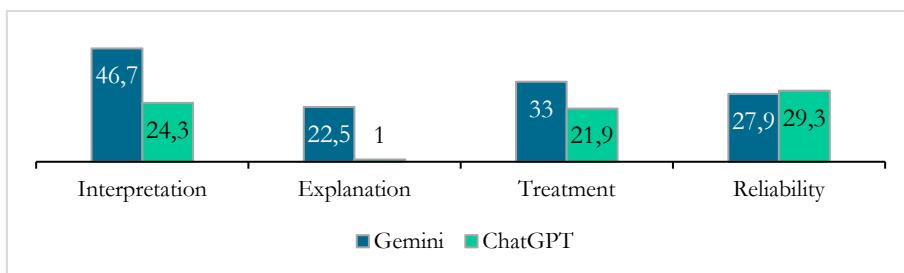

Figure 1: FRE scores for our pre-defined prompt themes

The Flesch Kincaid Reading Ease (FRE) score, a tool used to assess text readability on a scale from 0 to 100, where higher scores denote easier reading, was employed to evaluate the outputs from ChatGPT and Gemini across four different prompts. The analysis revealed that Gemini generally provided more accessible text than ChatGPT did. In the 'Interpretation' category, Gemini scored 46.7, which was markedly higher than ChatGPT's 24.3, indicating a clearer articulation of information. Similarly, in the 'Explanation' prompt, while both scores suggested difficulty, Gemini's 22.5 was significantly more comprehensible than ChatGPT's

notably low 1.0. In assessing 'Treatment' options, Gemini again outperformed with a score of 33 against ChatGPT's 21.9, underscoring a consistent trend in readability. The 'Reliability' scores were more comparable, with ChatGPT scoring 29.3 and Gemini 27.9, both indicating challenging texts yet showing slight ease with ChatGPT. This evaluation underscores a consistent pattern in which Gemini's outputs are easier to understand than those of ChatGPT, thus enhancing its utility in delivering patient education.

For **qualitative evaluation**, we defined five open questions to obtain direct feedback from the patient following the use of both chatbots. The questions and their answers are also included in the supplementary file.

The feedback from the patient confirmed the quantitative results. The patient appreciated the clearer explanations and structured responses from Gemini, which helped demystify medical jargon more effectively than ChatGPT. Both tools motivated the patient to pursue further medical consultations, but Gemini was preferred for its use of trusted sources and ability to break down information.

4   Summary and next steps

Our study assessed the effectiveness of ChatGPT and Gemini in supporting a patient with rare conditions such as syringomyelia and hydrocephalus, focusing on providing clear, actionable, and reliable information.

Evaluation results indicate that Gemini excels in delivering accessible and actionable information through a structured approach, including detailed term-by-term explanations and a concluding "Outlook." In contrast, ChatGPT, while matching in content quality and organization, lacks in actionability and requires improvements in language simplification. The patient feedback was positive, particularly for Gemini, suggesting that Generative AI can empower patients by effectively bridging the information gap. However, it is important to acknowledge that these findings are based on the experience of a single patient, highlighting the need for further research. Future studies will involve a larger cohort of patients with these conditions to validate and expand the initial findings.

P. Kröckel: *Generative AI in Assisting Patients with Syringomyelia and Hydrocephalus: A Preliminary Comparison of Chatgpt and Gemini*

677

In conclusion, this study highlights the potential of AI tools in patient education and healthcare management of rare diseases. Future research should extend to more patients and incorporate diverse evaluation metrics, such as patient engagement and information retention rates, to fully assess the impact of AI in patient care. This work underscores the necessity of developing patient-centered AI tools that complement traditional healthcare methods, fostering a well-supported patient community.

References

Ayoub, N. F., Lee, Y.-J., Grimm, D., & Divi, V. (2023). Head-to-Head Comparison of ChatGPT Versus Google Search for Medical Knowledge Acquisition. Otolaryngology - Head and Neck Surgery. Advance online publication. https://doi.org/10.1002/ohn.465

Bernard, S., Cooke, T., Cole, T., Hachani, L., & Bernard, J. (2018). Quality and readability of online information about type 2 diabetes and nutrition. JAAPA, 31(11), 41–44. https://doi.org/10.1097/01.JAA.0000546481.02560.4e

Bristol, R. E. (2014). Hydrocephalus. In Encyclopedia of the Neurological Sciences (pp. 636–639). Elsevier. https://doi.org/10.1016/B978-0-12-385157-4.00753-3

Campbell, D. J., Estephan, L. E., Sina, E. M., Mastrolonardo, E. V., Alapati, R., Amin, D. R., & Cottrill, E. E. (2023). Evaluating ChatGPT Responses on Thyroid Nodules for Patient Education. Thyroid : Official Journal of the American Thyroid Association. Advance online publication. https://doi.org/10.1089/thy.2023.0491

Chiesa-Estomba, C. M., Lechien, J. R., Vaira, L. A., Brunet, A., Cammaroto, G., Mayo-Yanez, M., Sanchez-Barrueco, A., & Saga-Gutierrez, C. (2023). Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. European Archives of Oto-Rhino-Laryngology : Official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : Affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery. Advance online publication. https://doi.org/10.1007/s00405-023-08104-8

Coskun, B., Ocakoglu, G., Yetemen, M., & Kaygisiz, O. (2023). Can ChatGPT, an Artificial Intelligence Language Model, Provide Accurate and High-quality Patient Information on Prostate Cancer? Urology, 180, 35–58. https://doi.org/10.1016/j.urology.2023.05.040

Gabriel, J., Shafik, L., Alanbuki, A., & Larner, T. (2023). The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. International Urology and Nephrology, 55(11), 2717–2732. https://doi.org/10.1007/s11255-023-03729-4

Gordon, E. B., Towbin, A. J., Wingrove, P., Shafique, U., Haas, B., Kitts, A. B., Feldman, J., & Furlan, A. (2024). Enhancing Patient Communication With Chat-GPT in Radiology: Evaluating the Efficacy and Readability of Answers to Common Imaging-Related Questions. Journal of the American College of Radiology : JACR, 21(2), 353–359. https://doi.org/10.1016/j.jacr.2023.09.011

Hydrocephalus Association. (2024). Shunt Systems. https://www.hydroassoc.org/shunt-systems/

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel. Chief of Naval Technical Training: Naval Air Station Memphis. Research Branch Report.

https://stars.library.ucf.edu/istlibrary/56?utm_source=stars.library.ucf.edu%2Fistlibrary%2F
56&utm_medium=PDF&utm_campaign=PDFCoverPages

Litzkendorf, S., Babac, A., Rosenfeldt, D., Schauer, F., Hartz, T., Lührs, V., Schulenburg, J.-M. G. v.,
& Frank, M. (2016). Information Needs of People with Rare Diseases - What Information
Do Patients and their Relatives Require? Journal of Rare Disorders: Diagnosis & Therapy,
2(2:40). https://raredisorders.imedpub.com/articles/information-needs-of-people-with-rare-
diseases--what-information-do-patients-and-their-relatives-require.pdf

Long, J. C., Best, S., Nic Giolla Easpaig, B., Hatem, S., Fehlberg, Z., Christodoulou, J., & Braithwaite,
J. (2022). Needs of people with rare diseases that can be supported by electronic resources: A
scoping review. BMJ Open, 12(9), e060394. https://doi.org/10.1136/bmjopen-2021-060394

McCarthy, C. J., Berkowitz, S., Ramalingam, V., & Ahmed, M. (2023). Evaluation of an Artificial
Intelligence Chatbot for Delivery of IR Patient Education Material: A Comparison with
Societal Website Content. Journal of Vascular and Interventional Radiology : JVIR, 34(10),
1760-1768.e32. https://doi.org/10.1016/j.jvir.2023.05.037

Milhorat, T. H. (2000). Classification of syringomyelia. Neurosurgical Focus, 8(3).
https://doi.org/10.3171/foc.2000.8.3.1

Pate, J. W., Tran, E., Radhakrishnan, S., & Leaver, A. M. (2021). The Importance of Perceived
Relevance: A Qualitative Evaluation of Patient's Perceptions of Value and Impact Following
a Low-Intensity Group-Based Pain Management Program. Medicina (Kaunas, Lithuania),
57(1). https://doi.org/10.3390/medicina57010046

Stanarević Katavić, S. (2019). Health information behaviour of rare disease patients: Seeking, finding
and sharing health information. Health Information and Libraries Journal, 36(4), 341–356.
https://doi.org/10.1111/hir.12261

Szmuda, T., Özdemir, C., Ali, S., Singh, A., Syed, M. T., & Słoniewski, P. (2020). Readability of online
patient education material for the novel coronavirus disease (COVID-19): A cross-sectional
health literacy study. Public Health, 185, 21–25. https://doi.org/10.1016/j.puhe.2020.05.041

WebFX. Readability Test [Computer software]. https://www.webfx.com/tools/read-able/

Zheng, Y.-C., Liu, Y.-T., Wei, K.-C., Huang, Y.-C., Chen, P.-Y., Hsu, Y.-H., & Lin, C.-L. (2023).
Outcome predictors and clinical presentation of syringomyelia. Asian Journal of Surgery,
46(2), 705–711. https://doi.org/10.1016/j.asjsur.2022.06.150