



Large language models: Are artificial intelligence-based chatbots a reliable source of patient information for spinal surgery?

Anna Stroop¹ · Tabea Stroop¹ · Samer Zawy Alsofy^{1,2} · Makoto Nakamura³ · Frank Möllmann⁴ · Christoph Greiner⁴ · Ralf Stroop^{1,5} 

Received: 12 March 2023 / Revised: 31 August 2023 / Accepted: 25 September 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Purpose Large language models (LLM) have recently attracted attention because of their enormous performance. Based on artificial intelligence, LLM enable dialogic communication using quasi-natural language that approximates the quality of human communication. Thus, LLM could play an important role for patients to become informed. To evaluate the validity of an LLM in providing medical information, we used one of the first high-performance LLM (ChatGPT) on the clinical example of acute lumbar disc herniation (LDH).

Methods Twenty-four spinal surgeons experienced in LDH surgery directed questions to ChatGPT about the clinical picture of LDH from a patient's perspective. They evaluated the quality of ChatGPT responses and its potential use in medical communication. The responses were compared with the information content of a standard informed consent form.

Results ChatGPT provided good results in terms of comprehensibility, specificity, and satisfaction of responses and in terms of medical accuracy and completeness. ChatGPT was not able to provide all the information that was provided in the informed consent form, but did communicate information that was not listed there. In some cases, albeit minor, ChatGPT made medically inaccurate claims, such as listing kyphoplasty and vertebroplasty as surgical options for LDH.

Conclusion With the incipient use of artificial intelligence in communication, LLM will certainly become increasingly important to patients. Even if LLM are unlikely to play a role in clinical communication between physicians and patients at the moment, the opportunities—but also the risks—of this novel technology should be alertly monitored.

Keywords ChatGPT · Large language model · Patient information · Spinal surgery

Background

Artificial intelligence (AI) has long since found its way into our everyday lives, such as the emergency braking or parking assistance functions in vehicles, but also into medicine. Automatic image recognition systems support endoscopic diagnostics just as AI enables big data analysis [1–3]. But we also listen and respond to AI in everyday communication, like the voice-assisted systems Alexa or Siri [3]. In social media, so-called chatbots (a portmanteau of *to chat* and *robot*) represent large language models (LLM) that—based on AI—enable dialogic communication using quasi-natural language that approximates the quality of interpersonal communication [4].

One of the first high-performing and freeware LLM, ChatGPT (generative pre-trained transformer, a performant chatbot from the company OpenAI), has been made available to the public for evaluation. Based on the machine-learning

✉ Ralf Stroop
ralf.stroop@uni-wh.de

¹ Faculty of Health, Department of Medicine, Witten-Herdecke University, Alfred-Herrhausen-Straße 45, 58455 Witten, Germany

² Department of Neurosurgery, St. Barbara-Hospital, Academic Hospital of Westfälische Wilhelms-University Münster, Hamm, Germany

³ Department of Neurosurgery, Academic Hospital Köln-Merheim, Witten-Herdecke University, Cologne, Germany

⁴ Department for Neuro- and Spine Surgery, Niels Stensen Neuro Center, Osnabrück, Germany

⁵ Medical School Hamburg, Hamburg, Germany

paradigms of supervised learning and reinforcement learning, ChatGPT involved human trainers to improve its performance [5, 6]. If ChatGPT is asked about its own database from which it generates its responses, the software gives this self-referential answer: “*ChatGPT was trained on text data available on the internet up until 2021. This includes a variety of sources, such as news articles, scientific documents, social media posts, and more. The exact size of the database is not publicly known, but it is estimated to be several terabytes of text.*” (openai.com).

In the hospital, physician communication with the patient is as important as it is time-consuming. Physician communication must be informative and empathetic in equal measure, and must be adapted to communicate at eye level with the patient. This is essential for taking the patient’s medical history and achieving informed consent for the required diagnostics and therapy—a sustainable patient compliance represents the relevant basis in the treatment prognosis [7–9].

Accordingly, we decided to examine the extent to which the chatbots rapidly entering the market could support medical communication with patients. For this purpose, we first evaluated the content validity of ChatGPT, with the results presented in this article.

Methods

To test the hypothesis that patients can become comprehensively informed about their medical situation using LLM, an online baseline survey was sent to 52 spinal surgeons experienced in the diagnosis, treatment, and at least microsurgery of lumbar disc herniation (LDH) in February 2023, using the web-based Unipark online survey software (<https://www.unipark.de/>, Tivian XI GmbH, Cologne, Germany). The survey included an orienting presentation of the ChatGPT chatbot, as well as registration instructions on the openai.com-homepage. The spinal surgeons were then given the task of imagining that they were a patient with an acute LDH resulting in very painful sciatica, but no sensorimotor deficit and no vegetative symptoms, and that surgery had been recommended. Study participants were instructed to inform themselves about the clinical picture of LDH, including symptoms, diagnosis, treatment options, and prognosis; it was important not to consider the information using their background of spinal surgical expertise, but as a layman patient suffering from sciatica.

Participants were asked to copy both the questions asked of ChatGPT and the answers provided by ChatGPT into the online questionnaire. They were also asked to rate the quality of the answers according to the categories presented in Online Resource 1. A maximum of 15 questions were evaluated within the survey tool. Participants were then asked to

provide some general evaluations of the use of ChatGPT in supportive patient information (Online Resource 2).

The information content of a standardized informed consent sheet (www.thieme.de), considered legally comprehensive, is assumed to represent the relevant knowledge for a patient suffering from an LDH. In the informed content sheet in the German language, 215 individual items of information were identified. These were assigned to six main categories: (1) clinical picture; (2) treatment options; (3) how is the operation performed?; (4) risks and possible associated complications; (5) what are the chances of success?; and (6) instruction advice. These were subclassified according to a four-digit key. Using the same key, the individual information was extracted from the answers given by ChatGPT. Percent coverage was determined by comparing the information from the ChatGPT responses in relation to the information from the informed consent sheet (Online Resource 3).

Data generated or analysed during this study are partially included in this published article and the Online Resources.

Results

The response rate to the questionnaire was 46% (24/52 surgeons). The respondents submitted a total of 139 questions to ChatGPT; two pairs of question/answer had to be discarded as they did not belong to the clinical picture of LDH. A median of four questions (min. 1, max. 15) were submitted per spinal surgeon.

The quality of the answers given by ChatGPT with respect to the clinical picture of LDH was rated as largely or even completely understandable by 97% of respondents (Fig. 1a). Overall, 97% of the ChatGPT responses came up with only some or no foreign words (Fig. 1b). In terms of how specifically the ChatGPT responses answered the questions, 31% of answers were rated as rather or much too general (Fig. 1c), while 86% were considered satisfactory (Fig. 1d); 44% were assessed as largely correct from a medical point of view, and 52% as completely correct (Fig. 1e). Overall, 55% of responses were considered medically complete and comprehensive (Fig. 1f). Against the background that ChatGPT enables dialogic communication using quasi-natural language that approximates the quality of interpersonal communication, it was investigated whether the response behaviour was empathetic. The communication was perceived as neutral in 82% of respondents, empathetic in 14%, and not very empathetic or even unsettling in 3% (Fig. 1g).

The overall evaluation of ChatGPT in medical communication, based on the example of the clinical picture of LDH, included the usability of the software; this was rated as intuitive by 100% of respondents. The registration process for creating the account was taken from this evaluation (Fig. 2a). Overall, 88% of respondents suspected that patients were

Fig. 1 Assessment of the quality of the answers provided by the ChatGPT for questions surrounding the clinical picture of a lumbar disc herniation

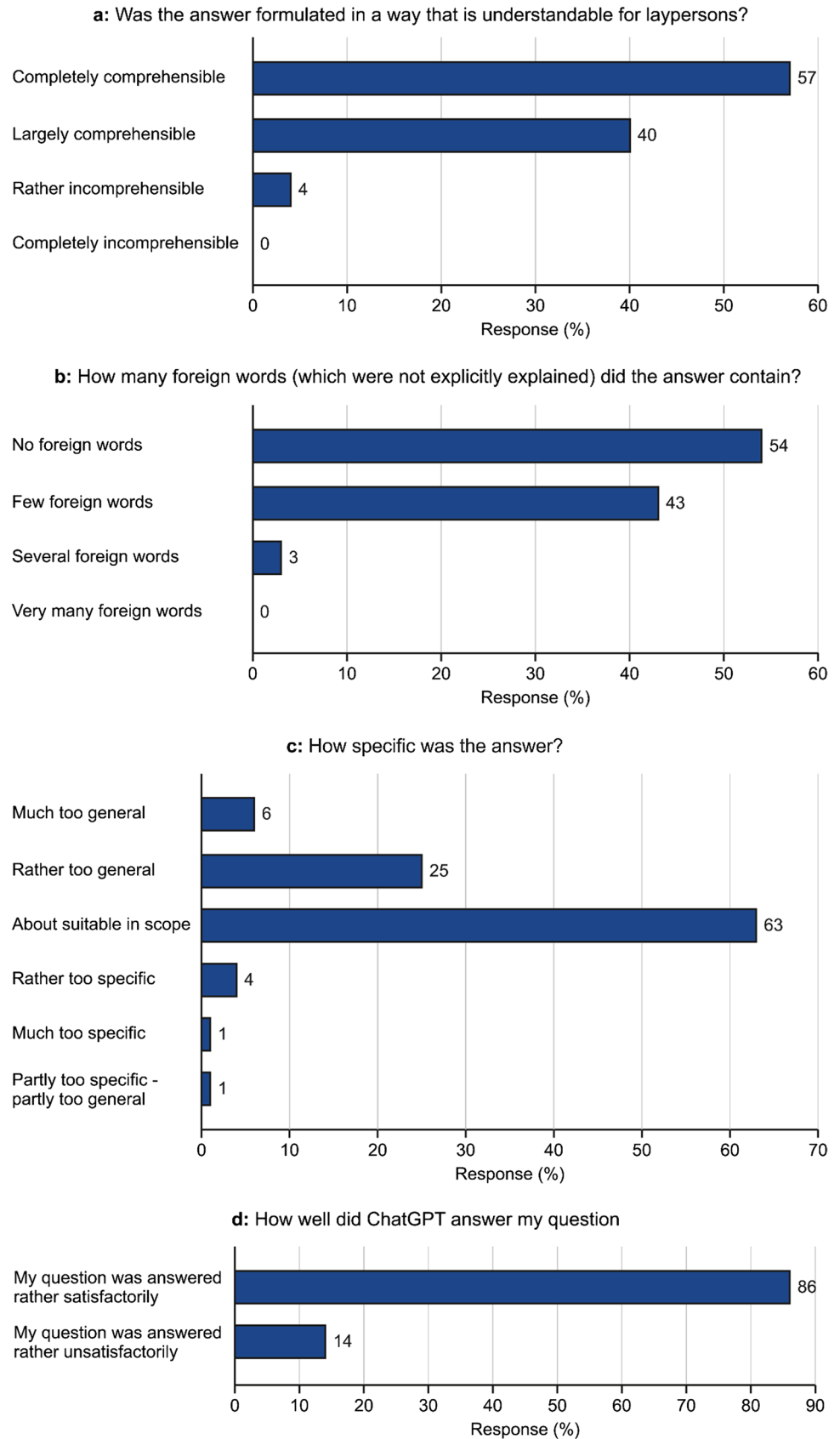
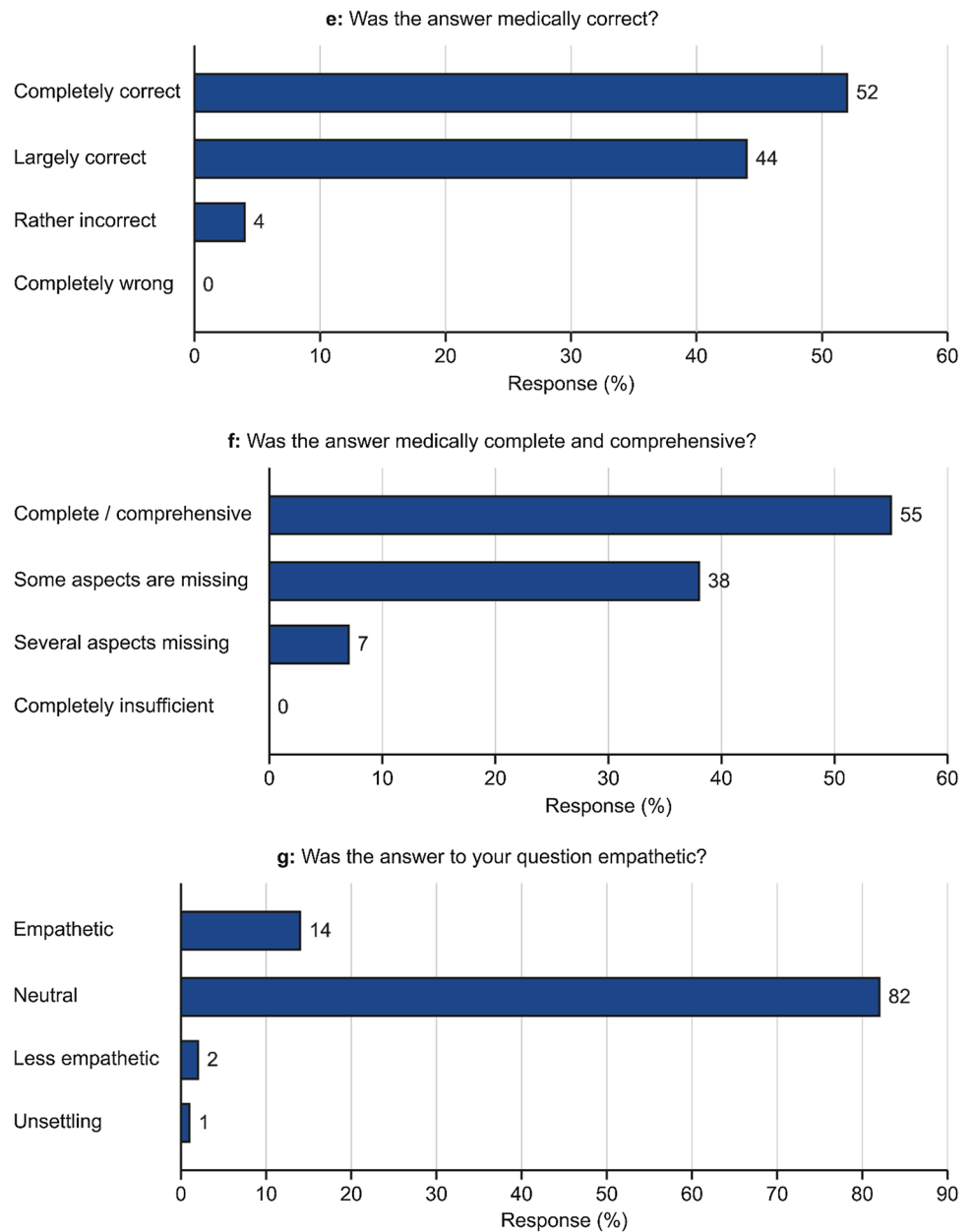


Fig. 1 (continued)

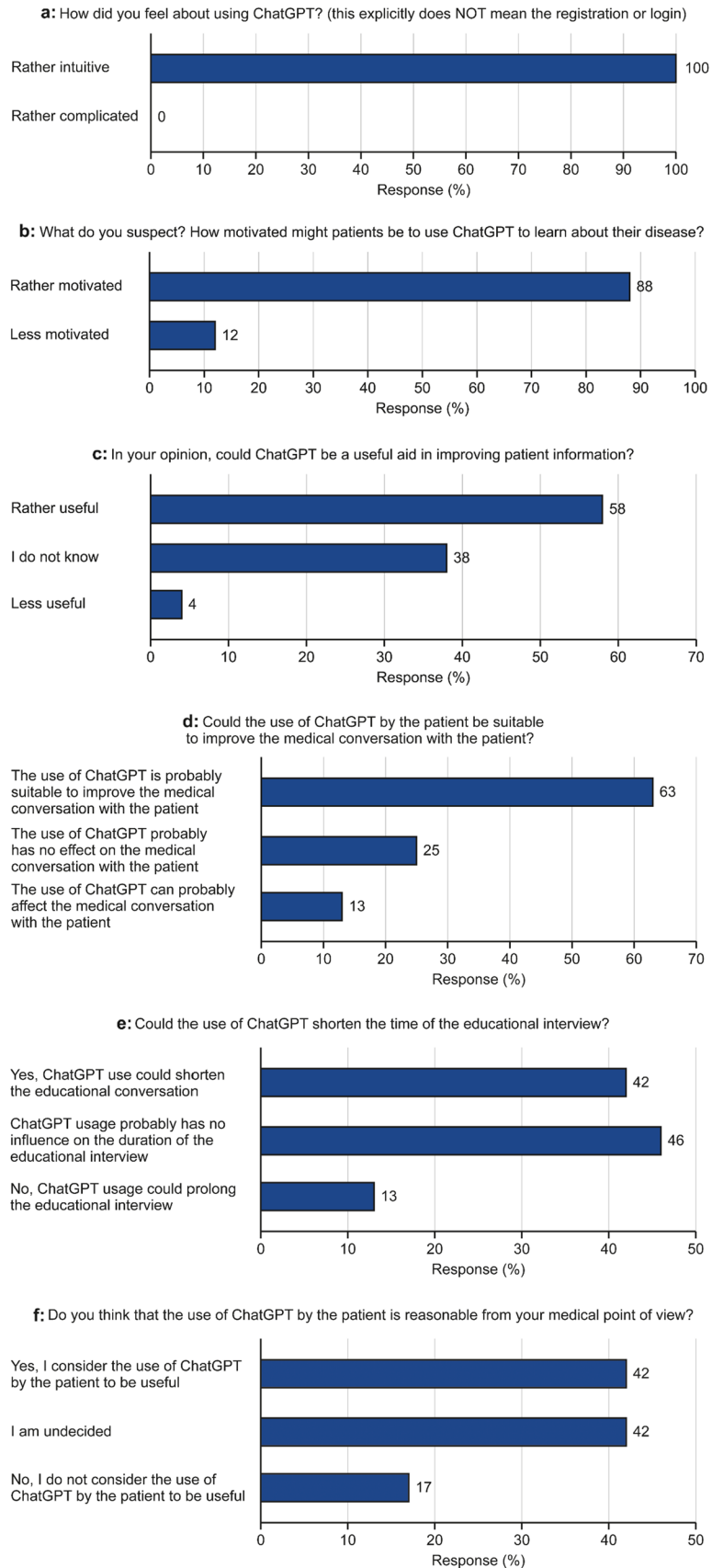


motivated to inform themselves about their clinical picture via ChatGPT (Fig. 2b); 58% agreed that ChatGPT could be a useful tool to improve patient information, 38% were undecided, and 4% considered ChatGPT to be less useful (Fig. 2c). Whether the use of ChatGPT by the patient could be suitable to improve the medical conversation between the patient and the doctor was rated as positive in 63% of respondents and indifferent in 25%, while 13% suspected an impairment in medical communication (Fig. 2d). It was assumed that the ChatGPT would shorten the time required for the patient's informed consent in 42% of respondents, whereas 46% suspected no effect and 13% even thought that informed consent might be prolonged (Fig. 2e). Use of the ChatGPT from a medical point of view was considered

useful in 42% of respondents, but 42% were undecided in this respect and 17% did not believe that use of the ChatGPT use was advisable (Fig. 2f).

The totality of all answers given by ChatGPT could be condensed to 151 distinguishable pieces of information. We compared the incidence of this information in relation to the information on the clinical picture of LDH contained in an informed consent form ($n = 215$). Of the information provided in the informed consent form, 48% ($n = 103$) was also covered by ChatGPT at least once. However, ChatGPT also provided information about the clinical picture of LDH that was not included in the consent form ($n = 48$ (22%) responses). The answers given by ChatGPT could be assigned to the corresponding six

Fig. 2 The overall evaluation of ChatGPT in medical communication, based on the example of the clinical picture of lumbar disc herniation



categories of the informed consent sheet, listed in Methods. In these individual categories, responses from ChatGPT that provided information from the consent form vs. instances where ChatGPT exceeded the consent form, respectively, were as follows (Fig. 3): (1) clinical picture,

$n = 10$ (32%) versus $n = 2$ (6%); (2) treatment options, $n = 5$ (50%) vs. $n = 8$ (80%); (3) how is the operation performed? $n = 45$ (42%) vs. $n = 8$ (7%); (4) risks and possible associated complications, $n = 12$ (48%) vs. $n = 3$ (12%); (5) what

Fig. 3 Frequency of occurrence of the information given by ChatGPT in relation to the information on the clinical picture of lumbar disc herniation contained in an informed consent form, assigned to six main categories: (1) clinical picture, (2) treatment options, (3) how is the operation performed?, (4) risks and possible associated complications, (5) what are the chances of success?, and (6) instruction advice. A total of 215 individual information items were identified in the informed consent form. The frequency with which the individual information was given in a total of 139 ChatGPT responses is shown. However, ChatGPT also provided information on the clinical picture of lumbar disc herniation that was not contained in the informed consent form. These information frequencies were inserted in the respective section next to the blank item. The textual description of the individual information can be found in the Online Resource 3



are the chances of success? $n = 21$ (84%) vs. $n = 16$ (64%); (6) instruction advice, $n = 10$ (59%) vs. $n = 11$ (65%).

In detail, ChatGPT provided information about the clinical picture of LDH that was not included in the consent form, such as the occurrence of sciatica, which is characteristic of the clinical picture. In addition, the symptoms of LDH were described in more detail. Compared to the informed consent form, the ChatGPT answers to the treatment options were more complete than the questions on risks and complications. Furthermore, ChatGPT mentioned some other therapy options, such as acupuncture or neural therapy, as a conservative treatment option. Pain therapy was specified with the naming of substance classes. The description of drug, physical, or physiotherapeutic treatment options was more comprehensive, as was that for possible postoperative complications, information on recovery and rehabilitation, and comments on postoperative quality of life and lifestyle adjustments. The textual description of the individual information can be found in the Online Resource 3.

ChatGPT also provided erroneous information, albeit only in isolated cases; the rate of incorrect answers generated by ChatGPT was numerically low at 2/151 (1.3%). However, this erroneous information directly concerned the choice of treatment options and can therefore be described as serious. Concretely, kyphoplasty and vertebroplasty were suggested as treatment options for herniated discs. Among the responses provided by ChatGPT that exceeded the content of the informed consent sheet, no response was classified as superfluous.

Discussion

The recently published chatbot “ChatGPT”—one of the first, high-performance LLM—is an AI-based dialogic communication system that uses quasi-natural language approximating the quality of interpersonal communication. We evaluated its use with respect to its validity in medical patient communication.

We found that the chatbot provided answers with a high level of general comprehensibility, even for laypeople, without the excessive use of foreign words or special medical expressions. Sometimes the answers were too general and unspecific, but from a medical point of view were largely formulated correctly and comprehensively. The intuitive usability of ChatGPT, combined with a presumably high motivation in patients to inform themselves using this tool, indicates that ChatGPT will become increasingly important in the transfer of medical information. It is important to state, however, that in Germany (and most other countries), ChatGPT cannot and must not play a legal role in obtaining informed consent from patients before surgery; according

to the German Civil Code, the provision of information is always a medical task that cannot be delegated [10].

In a few cases, ChatGPT consistently provided false statements. For example, the question “What other therapy can I do for a herniated disc?” was answered as follows: “Minimally invasive procedures for the treatment of lumbar disc herniation include: Endoscopic discectomy, ..., microdiscectomy, stabilization procedures such as *kyphoplasty* or *vertebroplasty* ...”. The latter two procedures were incorrect, demonstrating that ChatGPT cannot always stand up to medical scrutiny. It should be noted that ChatGPT does not claim to be medically correct when it provides an answer, but always refers to the necessity for patients to consult their physician for further information. Nevertheless, this means that while ChatGPT offers an opportunity for patients to become self-informed about their clinical picture, it also represents a risk. Even though ChatGPT has now been overtaken in terms of content by more powerful LLM, ChatGPT will still be important as a free source of information compared to paid-for offerings such as GPT-4.

It is well known that patients receive information from their treating physician, but that they also consult other sources of information, such as books, health guides, and online libraries, conversations with friends and family, and patient organizations. Online sources including social networks are becoming increasingly important [11, 12]. The internet offers a wealth of information on health topics, contained on medical websites, forums, and blogs where patients can ask questions and share information. However, not all online sources are reliable, which poses challenges in the use of social media for health purposes [13].

Online sources for patient information are often very well prepared graphically, with explanatory images or even videos. The lack of a graphical presentation was considered to be a distinct disadvantage of ChatGPT in our respondents, which is based exclusively on written language without any accompanying pictures, graphics, or videos. Considering the momentum with which ChatGPT was launched in November 2022, it is expected that linking to a voice assistant, integrating or referencing images or videos, could be the next step, as has already been piloted through interfaces with other programs [14, 15].

According to the ratings of the spinal surgeons in our survey, it appeared that ChatGPT could mostly provide comprehensive answers, but with a tendency towards more general statements. This means that the patient must already have a certain idea what to ask ChatGPT. In some cases, the patient may also have to repeat the questions in a more specific way; otherwise, they will not be able to obtain the same amount of information that is offered in the preformed patient informed consent forms. This effect was quite evident in the limited coverage of information from the ChatGPT responses in relation to the informed consent sheet,

particularly about risks and complications. On the other hand, ChatGPT provided information that was not included in the consent form; the most obvious example is that the consent form mentioned back pain but not sciatica as a characteristic consequence of a herniated disc, information that ChatGPT regularly provided.

As with some online (non-scientific) sources, the fact that ChatGPT does not reveal its sources of information, so the patient cannot critically review the references for themselves, is generally regarded as a critical problem. “*The learning algorithm of ChatGPT includes unsupervised and supervised learning. During the training process, the model was fed with a massive amount of text data, such as books, articles, and web pages, using unsupervised learning techniques to learn the underlying patterns and structure of language. To improve the accuracy and quality of the model's responses, it was also fine-tuned with a smaller set of labeled data, which were manually annotated and labeled by human experts. This process is known as supervised learning, where the model is trained to predict the correct output based on the input and the labeled data.*” (openai.com, ChatGPT-request onto the question: “*Is there a human supervised algorithm in learning for ChatGPT?*”). Thus, there is a possibility that medical information could be biased on a large scale by human influence. More broadly, and focused on medical issues in Public Health, this highlights the problem that the ability of chatbots to rapidly produce massive amounts of text could lead to the spread of misinformation on an unprecedented scale—resulting in an “AI-driven infodemic” as a new threat to public health [16].

In terms of information technology, a distinction can be made between statistical and neural language models. A statistical language model calculates the probability of predicting a word sequence based on a number of previous words (word history). The more powerful neural language models calculate the word context using neural networks (word vectors) based on the parameter settings. Although not the first, ChatGPT was the neuronal network-based LLM that attracted the most widespread public attention. The learning ability of LLM is determined, among other things, by the number of parameters. This parameter number determines how many nuances can be mapped from the learning data set of the model. It has been assumed that the number of parameters has increased from 175 billion for ChatGPT to at least 100 trillion for GPT-4 [17].

In addition to ChatGPT and GPT-4, offered by OpenAI, there are other commercial, enterprise-ready service providers. Alphabet, a company associated with Google, presented Bard as an AI-based chatbot in May 2023 [18]. Aleph Alpha's Luminous model, a European provider, presented in April 2023 an LLM that was said to be twice as powerful as ChatGPT in a benchmark test [19]. In the Asian region, Hyperclova was launched by the South Korean company

Naver in June this year, based on 204 billion parameters [20]. The rapid and worldwide release of new chatbots reveals the dynamic momentum and associated market expectations in the field of LLM.

In a medical application study, Ayers et al. investigated whether an AI chatbot assistant could provide answers to patient questions that were of comparable quality and empathy to those written by physicians. They found that 78.6% of health professionals preferred chatbot responses to physician responses. Furthermore, the chatbot responses were rated as being of significantly higher quality and more empathetic than the physician responses. The study concluded that further exploration of this technology was warranted in clinical settings, such as using chatbot to draft responses that physicians could then edit. Randomized trials were proposed to further investigate whether the use of AI assistants could improve responses as well as patient outcomes [21].

In the assessment by Au Yeung et al. [22], it is probable that conversational AI will soon be developed for use in healthcare, but is not yet ready for clinical use. On the one hand, this statement is derived from the comparison of ChatGPT and Foresight—an LLM that focuses on modeling patient data and disorders. The comparison was made on the task of forecasting relevant diagnoses based on clinical vignettes. Emphasizing the high-ranking patient safety and accuracy in the healthcare domain, they differentiated whether the tool was used by a clinician user (as clinical decision support) or by a patient (as an interactive medical chatbot). They considered the limitations of transformer-based chatbots for clinical use: the open internet database, on which OpenAI's ChatGPT is based, for example, brings potential limitations due to mirroring biases or lack of accurate detail. LLM that have been trained on biomedical data, such as BioMedLM as a domain-specific large language model for biomedical text [23], are subject to publishing trends rather than trends of actual patients and diseases in healthcare. Instead, the few LLM that are trained and validated on real-world clinical data due to sensitivity of patient data were emphasized [22]. The authors exemplify Gatron as a large clinical language model conceptualized to unlock patient information from unstructured electronic health records [24].

The information provided by AI in our study was limited to a circumscribed clinical picture from the physician's perspective and appraisal, albeit with them attempting to view the information from a patient's perspective. An evaluation of ChatGPT by patients and the efficiency of information provision needs to be performed in further studies—a requirement that was also emphasized in the literature regarding future use of AI chatbots in medicine [21, 22].

Finally, the problem of how far patients can and may be informed using AI systems remains an ethically important point of discussion. Thus, our study contributes to current

knowledge on the significance of chatbot-based communication in medicine. At present, LLM will not and must not replace medical communication between physicians and patients. But with 60 million visits per day for ChatGPT alone [25], the upcoming LLM will inevitably have a weighty role in the patient's own search for information. It is therefore important to perceive the possibilities of this AI-driven tool, but also the inherent problems associated with the software, which is not always error-free from a medical point of view.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00586-023-07975-z>.

Acknowledgements We thank Prof. Dr. habil. Mathias Kauff for assistance with the methodology and comments that greatly improved the manuscript. The first draft was reviewed by Deborah Nock (Medical WriteAway, Norwich, UK).

Funding The authors have no financial or proprietary interests in any material discussed in this article.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Yang YC, Islam SU, Noor A, Khan S, Afsar W, Nazir S (2021) Influential usage of big data and artificial intelligence in healthcare. *Comput Math Methods Med* 2021:1–13. <https://doi.org/10.1155/2021/5812499>
- Iqbal JD, Vinay R (2022) Are we ready for artificial intelligence in medicine? *Swiss Med Wkly* 152:w30179. <https://doi.org/10.4414/smw.2022.w30179>
- Mintz Y, Brodie R (2019) Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol* 28(2):73–81. <https://doi.org/10.1080/13645706.2019.1575882>
- Adamopoulou E, Moussiades L (2020) An overview of chatbot technology. In: Maglogiannis I, Iliadis L, Pimenidis E (eds) *Artificial intelligence applications and innovations*, vol 584. Springer International Publishing, Cham, pp 373–383
- Patel SB, Lam K (2023) ChatGPT: The future of discharge summaries? *The Lancet Digital Health* 5(3):e107–e108. [https://doi.org/10.1016/s2589-7500\(23\)00021-3](https://doi.org/10.1016/s2589-7500(23)00021-3)
- Aljanabi M, ChatGPT (2023) ChatGPT: future directions and open possibilities. *MJCS*. <https://doi.org/10.58496/mjcs/2023/003>
- Becker G, Kempf DE, Xander CJ, Momm F, Olschewski M, Blum HE (2010) Four minutes for a patient, twenty seconds for a relative - an observational study at a university hospital. *BMC Health Serv Res*. <https://doi.org/10.1186/1472-6963-10-94>
- Rothberg MB, Steele JR, Wheeler J, Arora A, Priya A, Lindenauner PK (2012) The relationship between time spent communicating and communication outcomes on a hospital medicine service. *J Gen Intern Med* 27(2):185–189. <https://doi.org/10.1007/s11606-011-1857-8>
- Barrier PA, Li JTC, Jensen NM (2003) Two words to improve physician-patient communication: what else? *Mayo Clin Proc* 78(2):211–214. <https://doi.org/10.4065/78.2.211>
- Grüneberg C, Ellenberger J, Götz I, Herrler S, Pückler R von, Retzlaff B, Siede W, Sprau H, Thorn K, Weidenkaff W, Weidlich D, Wicke H (2023) Bürgerliches Gesetzbuch. §§ 630c Abs. 2 S. 1, 630e Abs. 1, Abs. 2 S. 1 Nr. 1 BGB, 82., neubearbeitete Auflage. Beck'sche Kurz-Kommentare, Band 7. C.H. Beck, München
- Chung JE (2014) Social networking in online support groups for health: how online social networking benefits patients. *J Health Commun* 19(6):639–659. <https://doi.org/10.1080/10810730.2012.757396>
- McMullan M (2006) Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. *Patient Educ Couns* 63(1–2):24–28. <https://doi.org/10.1016/j.pec.2005.10.006>
- Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, Alsawas M, Majzoub AM, Benkhadra R, Seisa MO, Ding J, Prokop L, Murad MH (2019) Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. *J Gen Intern Med* 34(9):1884–1891. <https://doi.org/10.1007/s11606-019-05109-0>
- Shafeeg A, Shazhaev I, Mihaylov D, Tularov A, Shazhaev I (2023) Voice assistant integrated with chat GPT. *IJCS*. <https://doi.org/10.33022/ijcs.v12i1.3146>
- Wu C, Yin S, Qi W, Wang X, Tang Z, Duan N (2023) Visual ChatGPT: talking, drawing and editing with visual foundation models
- Angelis L de, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, Rizzo C (2023) ChatGPT and the rise of large language models: The new AI-driven infodemic threat in Public Health
- Wikipedia (2023) GPT-4. <https://en.wikipedia.org/wiki/GPT-4>. Accessed 07 Jul 2023
- Wikipedia Google Bard. https://de.wikipedia.org/wiki/Google_Bard. Accessed 07 Jul 2023
- Aleph Alpha (2023) Luminous Performance Benchmarks. <https://www.aleph-alpha.com/luminous-performance-benchmarks>. Accessed 07 Jul 2023
- Naver Corp. (2023) NAVER Unveils HyperCLOVA, Korea's First Hyperscale 'AI to Empower Everyone'. <https://www.naver.com/en/promotion/pressReleasesView/30686>. Accessed 07 Jul 2023
- Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM (2023) Comparing physician and artificial intelligence Chatbot Responses to patient questions posted to a public social media forum. *JAMA Intern Med* 183(6):589. <https://doi.org/10.1001/jamainternmed.2023.1838>
- Au Yeung J, Kraljevic Z, Luintel A, Balston A, Idowu E, Dobson RJ, Teo JT (2023) AI chatbots not yet ready for clinical use. *Front Digit Health*. <https://doi.org/10.3389/fdgh.2023.1161098>
- BioMedLM (2023): BioMedLM: a Domain-specific large language model for biomedical text. Online verfügbar unter <https://www.mosaicml.com/blog/introducing-pubmed-gpt>, zuletzt aktualisiert am 19.07.2023, zuletzt geprüft am 19.07.2023.
- Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, Compas C, Martin C, Flores MG, Zhang Y, Magoc T, Harle CA, Lipori G, Mitchell DA, Hogan WR, Shenkman EA, Bian J, Wu Y (2022) GatorTron: a large clinical language model to unlock patient information from unstructured electronic health records
- Nerdy Nav (2023) ChatGPT Statistics & User Numbers in July 2023. <https://nerdynav.com/chatgpt-statistics/>. Accessed 18 Jul 2023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.