



Article

Evaluation of AI ChatBots for the Creation of Patient-Informed Consent Sheets

Florian Jürgen Raimann ^{*}, Vanessa Neef, Marie Charlotte Hennighausen, Kai Zacharowski and Armin Niklas Flinspach

Department of Anaesthesiology, Intensive Care Medicine and Pain Therapy, Goethe University Frankfurt, University Hospital Frankfurt, Theodor-Stern Kai 7, 60590 Frankfurt, Germany; neef@med.uni-frankfurt.de (V.N.); m.hennighausen@med.uni-frankfurt.de (M.C.H.); zacharowski@med.uni-frankfurt.de (K.Z.); flinspach@med.uni-frankfurt.de (A.N.F.)

* Correspondence: raimann@med.uni-frankfurt.de; Tel.: +49 69-6301-5868; Fax: +49-69-6301-7695

Abstract: Introduction: Large language models (LLMs), such as ChatGPT, are a topic of major public interest, and their potential benefits and threats are a subject of discussion. The potential contribution of these models to health care is widely discussed. However, few studies to date have examined LLMs. For example, the potential use of LLMs in (individualized) informed consent remains unclear. **Methods:** We analyzed the performance of the LLMs ChatGPT 3.5, ChatGPT 4.0, and Gemini with regard to their ability to create an information sheet for six basic anesthesiologic procedures in response to corresponding questions. We performed multiple attempts to create forms for anesthesia and analyzed the results checklists based on existing standard sheets. **Results:** None of the LLMs tested were able to create a legally compliant information sheet for any basic anesthesiologic procedure. Overall, fewer than one-third of the risks, procedural descriptions, and preparations listed were covered by the LLMs. **Conclusions:** There are clear limitations of current LLMs in terms of practical application. Advantages in the generation of patient-adapted risk stratification within individual informed consent forms are not available at the moment, although the potential for further development is difficult to predict.

Keywords: informed consent; artificial intelligence; anesthesiology; large language model



Citation: Raimann, F.J.; Neef, V.; Hennighausen, M.C.; Zacharowski, K.; Flinspach, A.N. Evaluation of AI ChatBots for the Creation of Patient-Informed Consent Sheets. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1145–1153. <https://doi.org/10.3390/make6020053>

Academic Editors: Ahmad Taher Azar, Irena Spasić and Karin Verspoor

Received: 25 March 2024

Revised: 14 May 2024

Accepted: 22 May 2024

Published: 24 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the introduction of natural language processing (NLP) into the field of artificial intelligence (AI) and the market launch of the first freely available large language model (LLM) ChatGPT[®], there have been signs of increasing integration of AI in medical fields. Initial publications report a progressive replacement of humans in direct customer communication by AI systems, which is sometimes reflected in job cuts in this area [1–3]. It can be assumed that these developments in the private sector will also have an impact on the medical sector in the future [4–6]. The use of AI in the healthcare sector, particularly to improve doctor-patient communication, is the subject of intense discussion and research [7–9]. However, trends are emerging that support the use of AI in this area [10,11].

Whether its use makes sense at the present time depends on a number of factors.

It would be desirable if LLM could automate the process of informing patients and obtaining consent for medical procedures [12]. This could save cost-intensive personnel resources and allow a focus on core competencies [13]. By using NLP, information documents could be created individually and automatically. The optimization of existing information sheets is another option. In particular, linking to a patient's existing medical records with risk stratification represents opportunities for optimizing care using AI [14].

Although modern AI systems offer many advantages, they also have limitations, particularly with regard to the classification of specific risks and possible process risks [15]. The classification of specific medical risks often requires extensive knowledge of the context and

individual circumstances of the patient, the procedure, and the infrastructural environment. There is a danger that specific risks will be overlooked or misjudged and not mapped accordingly, so much so that even the World Health Organization (WHO) warns against the undifferentiated use of AI systems [16].

In today's society, providing legally compliant and individualized information prior to medical interventions is a core task of medical practice. To meet information requirements with a corresponding procedural mandate, numerous commercial providers make comfort forms available for corresponding interventions, sometimes even in multilingual and digital form [17]. Accordingly, there is an unmanageable liability risk when AI systems are used to create individualized information via LLM in the event of existing misinformation. This risk can be circumvented by standardized, legally verified forms [18].

Overall, it is currently unclear what potential exists for the use of AI systems to create accurate patient information documents for informed consent for medical interventions in anesthesia. Regardless of individual risk stratification, the extent to which currently available LLMs can be used to design a legally compliant patient information sheet for standard procedures has rarely been investigated [18,19]. The aim of this study was to evaluate the suitability of an LLM for the preparation of standardized patient information for consent for anesthesiology interventions from the end user's perspective without including patient-specific risks.

2. Materials and Methods

This study was conducted according to the ethical principles of the Helsinki Declaration (Ethical Principles for Medical Research Involving Human Subjects) [20]. Since the present study was only an evaluation of publicly available data created by an LLM and no clinical research was conducted with humans, there was no requirement for the involvement of an institutional review board (IRB; the ethics committee of the University Hospital Frankfurt). This manuscript adheres to the current CONSORT guidelines [21].

2.1. Preparation of the Prompts for LLM and Generation

As part of the preparation, a study team was developed that comprised four anesthesiologists with varying levels of expertise: one resident, one specialist, and two senior physicians. Two prompts were formulated for each of the six topics (general anesthesia, anesthesia for ambulatory surgery, peripheral regional anesthesia, peridural anesthesia, spinal anesthesia, and central venous line CVL) (see translated Supplement S1). For all prompts, the German language was used to compare German-based information sheets. These prompts were simultaneously fed to ChatGPT (version 3.5 and 4.0) and Google Gemini (7 July 2023 between 09:00 and 12:15 (CET)) to generate patient information. The selection was based on the fact that ChatGPT-4.0, as a more advanced and up-to-date (but currently charged) version of ChatGPT-3.5, may provide more accurate results. There is also a significant difference between ChatGPT-3.5 and Google Gemini in terms of access to the internet. Gemini has the ability to pull its answers from the internet in real-time, whereas ChatGPT-4 relies on a dataset that was current only until the end of 2021. This limitation means that ChatGPT-4 may not be able to provide the most up-to-date information, while Gemini can provide the latest answers to questions. In addition, there could be a significant difference due to the underlying training datasets of these LLMs. All responses were saved in individual Microsoft Word (MS Word 365, Microsoft Corporation, Redmond, WA, USA) documents. The results of the LLMs were prepared in German, analogous to the prompts, and were not translated during the follow-up.

2.2. Evaluation

The evaluation utilized commercially available questionnaires that are standard in the industry. These tools were chosen for their relevance and widespread use in assessing informed consent procedures. Diomed[®] questionnaires (Thieme Group, New York, NY, USA), and Perimed[®] questionnaires (perimed Fachbuch Verlag Dr. med. Straube GmbH, Fürth, Germany), in addition to questionnaires from a university hospital and district hospital were used. The use of the German language in the study was essential for comparing the AI-generated information sheets against standard German medical documents.

In the first step, a catalog of risks, complications, and relevant procedure descriptions was created for the official informed consent sheet ($n = 2$) and each of the topics ($n = 6$). In addition, five experts in the field of anesthesia were consulted in a concerted effort to review the checklists for completeness and, if necessary, to expand them based on existing guidelines for available informed consent items (procedures). A corresponding complete list of the items used in the checklists can be found in the translated Supplement S2.

In this process, the numbers of items to be investigated were identified, as shown in Table 1.

Table 1. Checklist items.

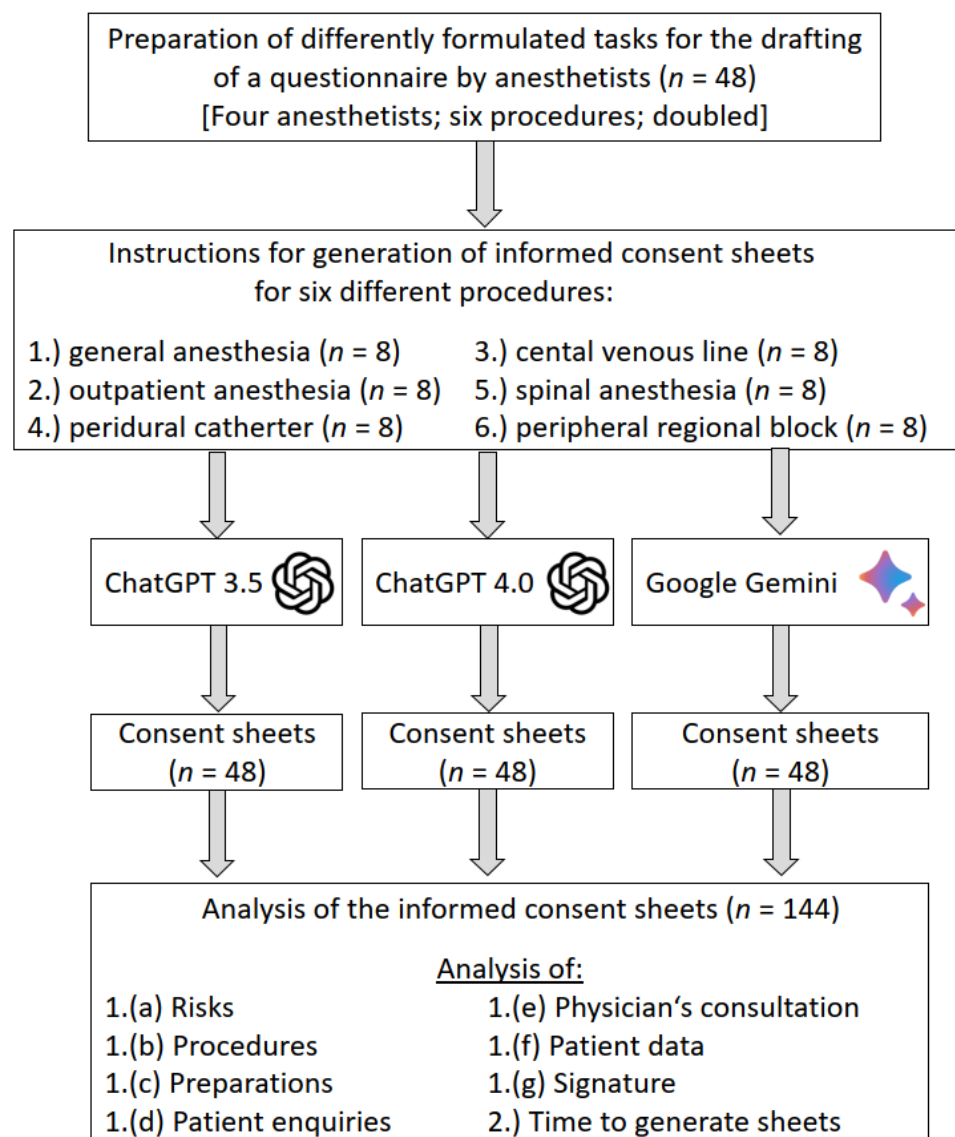
	General Anesthesia	Ambulant Anesthesia	Spinal Anesthesia	Epidural Catheter	Peripheral Nerve Block	Central Venous Line
Risks	25	3	25	28	21	24
Procedures	9	2	3	4	7	9
Preparations	5	8	2	4	3	1
Patient enquiries	1	1	1	1	1	1
Physicians' consultation	1	1	1	1	1	1
Patient data	1	1	1	1	1	1
Signature	1	1	1	1	1	1
Total	43	17	43	40	35	38

A tabular overview of the items concerted to the six selected anesthesiologic basis procedures with regard to the designated procedure-specific risk, procedure-specific information, preparation for the procedure, as well as the four general documentation items. These served as a reference for the sheets created by ChatGPT 3.5 & 4.0 and Google Gemini.

There were 216 items from 6 areas (including risk, procedure, preparations, and notice) across all six topics. This resulted in a total of 5328 data points with three LLM providers, a duplicate provision of a questionnaire, six topics, and 4 investigators.

In the second step, all automatically generated questionnaires were checked for congruence with the pre-generated catalogs. The agreement process was performed by two independent, experienced anesthesiologists. Disagreements led to the involvement of a third specialist, and a mutual agreement was reached to reach a consensus. The evaluation was conducted using a three-point scale. Possible answers were “applicable”, “not applicable”, and “paraphrase” (paraphrase of the characteristic in question).

Figure 1 shows the flow of the study including the numbers. The results were saved in a Microsoft Excel spreadsheet (Microsoft Corporation).



averaged percentage fulfillment levels of the LLMs is depicted in Figure 2. Figure 2 shows a comparison of the three LLMs used (green = Gemini, light blue ChatGPT 3.5., dark blue ChatGPT 4.0) with regard to the degree of fulfillment of the summarized aspects of legally compliant consent forms. The graphical representation is given as a percentage. Therefore, the circular structure at the outer edge would be expected for a fully satisfactory information form.

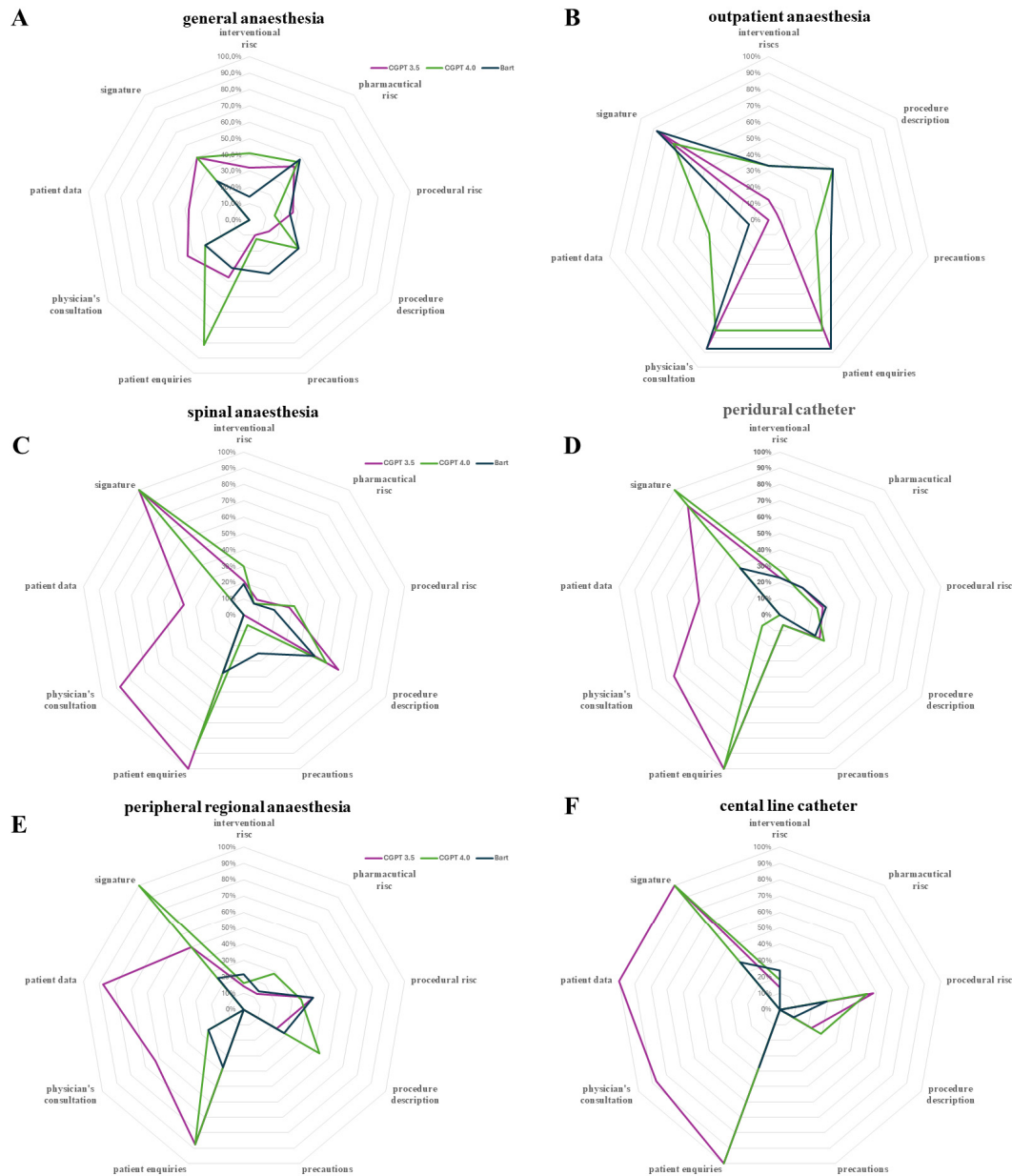


Figure 2. Grid diagram illustration of content item fulfillment. Grid diagram illustration of the level of fulfillment for the selected risk categories (where applicable, subdivided into procedural, pharmaceutical, and intervention-related risks), patient-oriented description of procedures, precautions, patient inquiries, physicians' consultation, patient data and assignment of a signature to the tested LLMs (ChatGPT 3.5, ChatGPT 4.0 and Gemini). Presentation of the percentage degree of fulfillment for the tested LLMs in relation to the designated category displayed at the corresponding edge. A: Regarding basic general anaesthesia, B: Regarding outpatient analgesia, C&D: For the neuraxial procedures, spinal anaesthesia, and epidural anaesthesia, E: For peripheral nerve blocks, F: For the insertion of a central venous catheter.

Certain formulations of the drafting of appropriate consent sheets did not lead to any results. Only rewording in the LLMs produced a corresponding draft document. Furthermore, there were significant differences in the time required to create the anesthesia consent sheets (Figure 3). Although the paid version of ChatGPT 4.0 promises faster creation, we were unable to detect this difference in our investigation. In contrast, even the processing time was significantly slower when using ChatGPT 4.0 than the free version of ChatGPT 3.5 ($p < 0.001$) and Google Gemini ($p < 0.001$). Across all three systems, Google Gemini was the fastest system at 12.1 s (± 2.1 s), followed by ChatGPT 3.5 (27.4 s (± 10.6 s)) and ChatGPT 4.0 (85.9 s (± 9.9 s)). It took Google Gemini 09:42 min, ChatGPT 3.5 21:53 min, and ChatGPT 4.0 68:43 min to create all of the questionnaires.

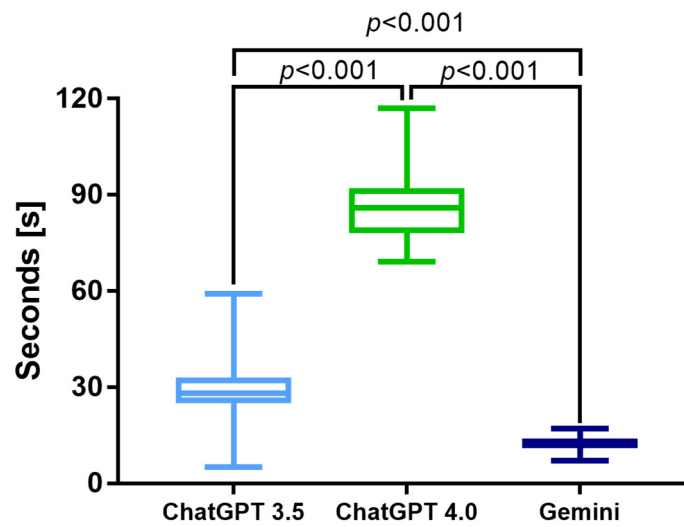


Figure 3. Boxplot visualization of the time required by the LLMs to carry out the task. Diverging completion times for patient-informed consent sheets among the different LLMs. Visualization of all 48 search queries per LLM in a boxplot.

4. Discussion

The results of our study show that LLMs do not currently appear to be able to create patient information sheets for standard anesthesiology procedures and, therefore, do not currently represent an alternative to commercially available products.

The use of artificial intelligence in the field of medicine has attracted great interest and considerable concern. The purpose of this study was to test whether popular LLMs that are popular among medical laypersons and physicians are capable of generating legally compliant patient consent forms. However, it was explicitly not the intention of this study to investigate the most suitable AI tool for this task or to test machine learning approaches with regard to automated document analysis. For barely six months, the major medical journals have been asking numerous critical questions on special issues about patient safety and the difficulties that may arise for physicians when using AI [22,23]. The first discussion addressed issues related to regulating language models such as ChatGPT [24]. The potential solutions offered by artificial intelligence appear to be highly promising, particularly in the analysis of constantly growing datasets in various areas of medical fields [25]. However, due to LLMs' applicability to the public and comprehensibility for laypeople, they are indisputably beyond the scope of national regulations or the wishes of professional societies with regard to the decision to apply them in medical settings. Accordingly, at present, there are mainly descriptive attempts to make this new technology understandable for medical professionals and to critically classify the results of its application.

With regard to the further development of patient-specific models of education, numerous obstacles have been discussed by Hunter and Holmes. The authors identified poor verifiability of automated evidence in terms of the selection and quality classification in the previous literature, as well as related risks [26]. However, our comparison of the two

versions of ChatGPT versions (3.5 without online data access and 4.0) indicates that little incorporation of the existing literature has occurred. Therefore, it remains unclear how an AI determines a corresponding risk allocation for a specific anesthesia procedure or a specific intervention; it may thus be subject to severe errors. The considerable difficulty for LLMs in generating legally compliant patient information could be due to the small number of questionnaires available online, most of which are subject to strict copyright. Similarly, the risks are not based on any clear logic but, for example, on past case law. In addition to information on standardized anesthesia procedures, an algorithm for individualized consent forms would require a risk assessment, which is difficult to obtain from the complex specialist literature and requires a correspondingly detailed evaluation by a specialist. A corresponding proof of concept should be manually checked on a corresponding dataset, but forensic hazards would still be possible in the event of a secondary misjudgment resulting in patient harm [27,28].

There are few relevant differences between the three large language models tested. This appears all the more astonishing in a direct comparison of ChatGPT 3.5 with ChatGPT 4.0; the commercial version 4.0 was supposed to include online data in the generative answers, although this version did not deliver quantifiable advantages. Nonetheless, numerous informed consent sheets from various clinics can be found online under common keywords. In some cases, even lists of relevant content for informed consent sheets provided by specialist societies can be found. In this respect, the results we generated may be interpreted as indicating that the training datasets are primarily responsible for the answers to our various queries.

The performance of the largest LLMs that are available online for free creation of patient information sheets for the most common anesthesiologic procedures, which we examined as an example, illustrates their inadequate application in medical specialist settings. The models fail with a consistently inadequate identification of the relevant risks and inadequate patient education. It was unlikely that the models would prove to be fully suitable immediately after their creation. However, the extent of the pending issues suggested a frightening current situation in the context of the general public's access to these models and their expected private application to medical questions by laypersons.

The initial phrasing of the task requesting the creation of patient informed consent sheets for patients appeared to have surprisingly limited influence on the results. The LLM revealed a remarkably consistent task completion in this respect, which can be linked to the operational principles of the model. However, in individual cases, language models were unable to generate the form based on the phrasing due to unclear circumstances. Furthermore, the significant differences in the time needed to complete the task were surprising. Although the ChatGPT application simulates a display similar to the manual data entry of typing when a task is presented, the fact that the commercial version took more than 45 s on average, which was almost twice as long as the other model, is difficult to explain. The supposed parallel online research to answer the question did not provide any significant added value in terms of content. However, it remains unclear how access to the internet affects the compatibility of LLMs. Further studies should conduct repeated assessments of performance.

5. Limitations

The performance of an LLM is based in part on the database with which the systems are trained in advance. The underlying evaluation is based on expert consensus and, therefore, represents a further bias, as no objectifiable parameters can be used for analysis, which reduces the informative value. This study is the first of its kind and thus provides an initial insight into the performance of these systems in a specific medical application. Depending on the local requirements for the legally binding nature of information sheets, their applicability may be assessed differently. The correct use of AI/ML and LLM is one of the decisive weak points. As discussed previously, inaccurate training datasets can have a significant impact on the completeness of the results and the probability of correct

predictions. For the undifferentiated use of LLM by the lay user, the accuracy of the training dataset for the corresponding questions cannot be verified [29]. The use of the German language in this study may limit the generalizability of the findings to other linguistic contexts. However, it was necessary to compare results with standard German medical documents. Therefore, further investigations with the aid of applications specially trained for this use case are necessary.

6. Conclusions

Our study exemplifies the clear limitations of current large language models in terms of practical application and the corresponding lack of compatibility with routine medical situations. The potential advantages in the generation of patient-adapted risk stratification within informed consent sheets do, therefore, not appear to be foreseeable, but the huge potential for future development is difficult to predict.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/make6020053/s1>.

Author Contributions: A.N.F.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing. F.J.R.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing. K.Z.: Methodology, Resources, Validation, Writing—review & editing. M.C.H.: Data curation, Formal analysis, Investigation, Validation, Writing—review & editing. V.N.: Formal analysis, Resources, Validation, Writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded solely by institutional resources.

Data Availability Statement: Data and further information are available upon request by the corresponding author.

Conflicts of Interest: A.N.F. received speaker fees from P.J. Dahlhausen & Co. GmbH, Cologne, Germany, and received the Sedana Medical Research Grant 2020 and the Thieme Teaching Award 2022. F.J.R. received speaker fees from Helios Germany, University Hospital Würzburg, and Keller Medical GmbH. FJR received financial support from HemoSonics LLC, pharma-consult Petersohn, and Boehringer Ingelheim. K.Z. has received honoraria for participation in advisory board meetings for Haemonetics and Vifor and received speaker fees from CSL Behring, Masimo, Pharmacosmos, Boston Scientific, Salus, iSEP, Edwards, and GE Healthcare. He is the Principal Investigator of the EU-Horizon 2020 project ENVISION (Intelligent plug-and-play digital tool for real-time surveillance of COVID-19 patients and smart decision-making in Intensive Care Units) and Horizon Europe 2021 project COvend (Biomarker and AI-supported FX06 therapy to prevent progression from mild and moderate to severe stages of COVID-19). The Department of Anesthesiology, Intensive Care Medicine & Pain Therapy of the University Hospital Frankfurt, Goethe University, received support from B. Braun Melsungen, CSL Behring, Fresenius Kabi, and Vifor Pharma for the implementation of Frankfurt's Patient Blood Management program. V.N. received speaker fees from Sysmex and Pharmacosmos. M.C.H. declared that there are no conflicts of interest.

References

1. Verma, P.; De Vynck, G. ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners. The Washington Post. Available online: <https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/> (accessed on 4 March 2024).
2. Cerullo, M. Here's How Many U.S. Workers ChatGPT Says It Could Replace. 2023. Available online: <https://www.cbsnews.com/news/chatgpt-artificial-intelligence-jobs/> (accessed on 26 June 2023).
3. Tangalakis-Lippert, K. IBM Halts Hiring for 7,800 Jobs That Could Be Replaced by AI, Bloomberg Reports. 2023. Available online: <https://www.businessinsider.com/ibm-halts-hiring-for-7800-jobs-that-could-be-replaced-by-ai-report-2023-5> (accessed on 26 June 2023).
4. Andriola, C.; Ellis, R.P.; Siracuse, J.J.; Hoagland, A.; Kuo, T.C.; Hsu, H.E.; Walkey, A.; Lasser, K.E.; Ash, A.S. A Novel Machine Learning Algorithm for Creating Risk-Adjusted Payment Formulas. *JAMA Health Forum* **2024**, *5*, e240625. [CrossRef] [PubMed]
5. Newman-Toker, D.E.; Sharfstein, J.M. The Role for Policy in AI-Assisted Medical Diagnosis. *JAMA Health Forum* **2024**, *5*, e241339. [CrossRef] [PubMed]

6. Tai-Seale, M.; Baxter, S.L.; Vaida, F.; Walker, A.; Sitapati, A.M.; Osborne, C.; Diaz, J.; Desai, N.; Webb, S.; Polston, G.; et al. AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication. *JAMA Netw. Open* **2024**, *7*, e246565. [[CrossRef](#)] [[PubMed](#)]
7. Mello, M.M.; Guha, N. ChatGPT and Physicians' Malpractice Risk. *JAMA Health Forum* **2023**, *4*, e231938. [[CrossRef](#)] [[PubMed](#)]
8. Hswen, Y.; Abbasi, J. AI Will—And Should—Change Medical School, Says Harvard's Dean for Medical Education. *JAMA* **2023**, *330*, 1820–1823. [[CrossRef](#)] [[PubMed](#)]
9. Medicine NEJo. Prescribing Large Language Models for Medicine: What's The Right Dose? NEJM Group. 2023. Available online: <https://events.nejm.org/events/617> (accessed on 2 November 2023).
10. Odom-Forren, J. The Role of ChatGPT in Perianesthesia Nursing. *J. PeriAnesthesia Nurs.* **2023**, *38*, 176–177. [[CrossRef](#)] [[PubMed](#)]
11. Neff, A.S.; Philipp, S. KI-Anwendungen: Konkrete Beispiele für den ärztlichen Alltag. *Deutsches Ärzteblatt* **2023**, *120*, A-236/B-207.
12. Anderer, S.; Hswen, Y. Will Generative AI Tools Improve Access to Reliable Health Information? *JAMA* **2024**, *331*, 1347–1349. [[CrossRef](#)] [[PubMed](#)]
13. Obradovich, N.; Johnson, T.; Paulus, M.P. Managerial and Organizational Challenges in the Age of AI. *JAMA Psychiatry* **2024**, *81*, 219–220. [[CrossRef](#)] [[PubMed](#)]
14. Sonntagbauer, M.; Haar, M.; Kluge, S. Künstliche Intelligenz: Wie werden ChatGPT und andere KI-Anwendungen unseren ärztlichen Alltag verändern? *Med. Klin.—Intensivmed. Notfallmedizin* **2023**, *118*, 366–371. [[CrossRef](#)]
15. Menz, B.D.; Modi, N.D.; Sorich, M.J.; Hopkins, A.M. Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance: Weapons of Mass Disinformation. *JAMA Intern. Med.* **2024**, *184*, 92–96. [[CrossRef](#)] [[PubMed](#)]
16. dpa. Weltgesundheits Organisation Warnt vor Risiken durch Künstliche Intelligenz im Gesundheitssektor. 2023. Available online: <https://www.aerzteblatt.de/treffer?mode=s&wo=1041&typ=1&nid=143259&s=ChatGPT> (accessed on 26 June 2023).
17. Lühnen, J.; Mühlhauser, I.; Steckelberg, A. The Quality of Informed Consent Forms. *Dtsch. Ärzteblatt Int.* **2018**, *115*, 377–383. [[CrossRef](#)]
18. Ali, R.; Connolly, I.D.; Tang, O.Y.; Mirza, F.N.; Johnston, B.; Abdulrazeq, H.F.; Galamaga, P.F.; Libby, T.J.; Sodha, N.R.; Groff, M.W.; et al. Bridging the literacy gap for surgical consents: An AI-human expert collaborative approach. *NPJ Digit. Med.* **2024**, *7*, 63. [[CrossRef](#)]
19. Mirza, F.N.; Tang, O.Y.; Connolly, I.D.; Abdulrazeq, H.A.; Lim, R.K.; Roye, G.D.; Priebe, C.; Chandler, C.; Libby, T.J.; Groff, M.W.; et al. Using ChatGPT to Facilitate Truly Informed Medical Consent. *NEJM AI* **2024**, *1*, A1cs2300145. [[CrossRef](#)]
20. PPR Human Experimentation. Code of ethics of the world medical association. Declaration of Helsinki. *Br. Med. J.* **1964**, *2*, 177. [[CrossRef](#)] [[PubMed](#)]
21. Schulz, K.F.; Altman, D.G.; Moher, D.; the CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Trials* **2010**, *11*, 32. [[CrossRef](#)] [[PubMed](#)]
22. Duffourc, M.; Gerke, S. Generative AI in Health Care and Liability Risks for Physicians and Safety Concerns for Patients. *JAMA* **2023**, *330*, 313. [[CrossRef](#)] [[PubMed](#)]
23. Kanter, G.P.; Packel, E.A. Health Care Privacy Risks of AI Chatbots. *JAMA* **2023**, *330*, 311–312. [[CrossRef](#)] [[PubMed](#)]
24. Minssen, T.; Vayena, E.; Cohen, I.G. The Challenges for Regulating Medical Use of ChatGPT and Other Large Language Models. *JAMA* **2023**, *330*, 315–316. [[CrossRef](#)] [[PubMed](#)]
25. Gomes, B.; Ashley, E.A. Artificial Intelligence in Molecular Medicine. *N. Engl. J. Med.* **2023**, *388*, 2456–2465. [[CrossRef](#)] [[PubMed](#)]
26. Hunter, D.J.; Holmes, C. Where Medical Statistics Meets Artificial Intelligence. *N. Engl. J. Med.* **2023**, *389*, 1211–1219. [[CrossRef](#)] [[PubMed](#)]
27. Wachter, R.M.; Brynjolfsson, E. Will Generative Artificial Intelligence Deliver on Its Promise in Health Care? *JAMA* **2024**, *331*, 65–69. [[CrossRef](#)] [[PubMed](#)]
28. Yalamanchili, A.; Sengupta, B.; Song, J.; Lim, S.; Thomas, T.O.; Mittal, B.B.; Abazeed, M.E.; Teo, P.T. Quality of Large Language Model Responses to Radiation Oncology Patient Care Questions. *JAMA Netw. Open* **2024**, *7*, e244630. [[CrossRef](#)] [[PubMed](#)]
29. Rocchetti, M.; Delnevo, G.; Casini, L.; Salomoni, P. A Cautionary Tale for Machine Learning Design: Why We Still Need Human-Assisted Big Data Analysis. *Mob. Netw. Appl.* **2020**, *25*, 1075–1083. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.