Journal of Cranio-Maxillo-Facial Surgery xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Cranio-Maxillo-Facial Surgery

journal homepage: www.jcmfs.com

Evaluating AI-Generated informed consent documents in oral surgery: A comparative study of ChatGPT-4, Bard gemini advanced, and human-written consents

Luigi Angelo Vaira [a,b,*,1], Jerome R. Lechien [c,d,1], Antonino Maniaci [e], Giuseppe Tanda [f], Vincenzo Abbate [g], Fabiana Allevi [h], Antonio Arena [g], Giada Anna Beltramini [i,l], Michela Bergonzani [m], Alessandro Remigio Bolzoni [l], Salvatore Crimi [n], Andrea Frosolini [o], Guido Gabriele [o], Fabio Maglitto [p], Miguel Mayo-Yáñez [q], Ludovica Orrù [f], Marzia Petrocelli [r], Resi Pucci [s], Alberto Maria Saibene [t], Stefania Troise [g], Alessandro Tel [u], Valentino Vellone [v], Carlos Miguel Chiesa-Estomba [w], Paolo Boscolo-Rizzo [x], Giovanni Salzano [g,p,2], Giacomo De Riu [a,2]

[a] Maxillofacial Surgery Operative Unit, Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy
[b] PhD School of Biomedical Science, Biomedical Sciences Department, University of Sassari, Sassari, Italy
[c] Department of Anatomy and Experimental Oncology, Mons School of Medicine, UMONS. Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium
[d] Department of Otolaryngology-Head Neck Surgery, Elsan Polyclinic of Poitiers, Poitiers, France
[e] Department of Medicine and Surgery, University of Enna Kore, Enna, Italy
[f] Dental School, Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy
[g] Head and Neck Section, Department of Neurosciences, Reproductive and Odontostomatological Science, Federico II University of Naples, Naples, Italy
[h] Maxillofacial Surgery Department, ASSt Santi Paolo e Carlo, University of Milan, Milan, Italy
[i] Department of Biomedical, Surgical and Dental Sciences, University of Milan, Milan, Italy
[l] Maxillofacial and Dental Unit, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy
[m] Maxillo-Facial Surgery Division, Head and Neck Department, University Hospital of Parma, Parma, Italy
[n] Department of CHIRMED. Maxillofacial Surgery Section, University of Catania, Catania, Italy
[o] Maxillofacial Surgery Unit, Department of Medical Biotechnologies, University of Siena, Siena, Italy
[p] Maxillo-Facial Surgery Unit, University of Bari "Aldo Moro", Bari, Italy
[q] Otorhinolaryngology, Head and Neck Surgery Department, Complexo Hospitalario Universitario A Coruña (CHUAC), A Coruña, Galicia, Spain
[r] Maxillofacial Surgery Operative Unit, Bellaria and Maggiore Hospital, Bologna, Italy
[s] Maxillofacial Surgery Unit, San Camillo-Forlanini Hospital, Rome, Italy
[t] Otolaryngology Unit, Santi Paolo e Carlo Hospital, Department of Health Sciences, University of Milan, Milan, Italy
[u] Clinic of Maxillofacial Surgery, Department of Head & Neck Surgery and Neuroscience, University Hospital of Udine, Italy
[v] Maxillofacial Surgery Unit, "S. Maria" Hospital, Terni, Italy
[w] Department of Otorhinolaryngology-Head & Neck Surgery, Hospital Universitario Donostia, San Sebastian, Spain
[x] Department of Medical, Surgical and Health Sciences, Section of Otolaryngology, University of Trieste, Trieste, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Informed consent
Oral surgery
Maxillofacial surgery
Patient education
Document quality
Large language models

ABSTRACT

This study evaluates the quality and readability of informed consent documents generated by AI platforms ChatGPT-4 and Bard Gemini Advanced compared to those written by a first-year oral surgery resident for common oral surgery procedures. The evaluation, conducted by 18 experienced oral and maxillofacial surgeons, assessed consents for accuracy, completeness, readability, and overall quality.

ChatGPT-4 consistently outperformed both Bard and human-written consents. ChatGPT-4 consents had a median accuracy score of 4 [IQR 4-4], compared to Bard's 3 [IQR 3–4] and human's 4 [IQR 3–4]. Completeness scores were higher for ChatGPT-4 (4 [IQR 4–5]) than Bard (3 [IQR 3–4]) and human (4 [IQR 3–4]). Readability

* Corresponding author. xillofacial surgery operative unit, Viale San Pietro 43B, Italy.
E-mail address: lavaira@uniss.it (L.A. Vaira).
[1] These authors contributed equally to this work and should be considered as co-first authors.
[2] These authors contributed equally to this work and should be considered as co-senior authors.

https://doi.org/10.1016/j.jcms.2024.10.002
Received 24 June 2024; Received in revised form 2 October 2024; Accepted 3 October 2024

AI in healthcare
Consent accuracy

was also superior for ChatGPT-4, with a median score of 4 [IQR 4–5] compared to Bard and human consents, both at 4 [IQR 4-4] and 4 [IQR 3–4], respectively. The Gunning Fog Index for ChatGPT-4 was 17.2 [IQR 16.5–18.2], better than Bard's 23.1 [IQR 20.5–24.7] and the human consents' 20 [IQR 19.2–20.9].

Overall, ChatGPT-4's consents received the highest quality ratings, underscoring AI's potential in enhancing patient communication and the informed consent process. The study suggests AI can reduce misinformation risks and improve patient understanding, but continuous evaluation, oversight, and patient feedback integration are crucial to ensure the effectiveness and appropriateness of AI-generated content in clinical practice.

1. Introduction

Artificial intelligence (AI) platforms are increasingly recognized for their potential impact on healthcare communication, particularly in patient information dissemination (Liu et al., 2023). Research on AI's ability to provide health information has shown a variable degree of accuracy, highlighting the need for ongoing evaluation and cautious adoption of AI applications in healthcare (Ayers et al., 2023; Hopkins et al., 2023; Chiesa-Estomba et al., 2024; Lorenzi et al., 2024; Radulesco et al., 2024; Vaira et al., 2024).

Within the domain of patient information, informed consent is perhaps the most crucial aspect prior to surgical procedures, bearing significant ethical and legal implications (Kinnersley et al., 2013). It facilitates patient autonomy, enabling informed decision-making regarding surgical interventions. The integrity of informed consent documentation is crucial, requiring accuracy, completeness, and patient comprehension. These documents must comprehensively cover the procedure, including associated risks, benefits, and alternatives (Cocanour, 2017; Glaser et al., 2020).

This necessity has propelled the adoption of electronic informed consent (eConsent) as a standard practice. Traditional informed consent often involved paper-based documents, which presented limitations including physical storage issues, accessibility challenges for patients and healthcare providers, and difficulties in ensuring that all patients received the most up-to-date information. Additionally, the comprehension of complex medical terminology and the engagement of patients in the consent process were significant hurdles (Guarino et al., 2022). AI can play a pivotal role in the eConsent ecosystem by further personalizing the patient experience. AI algorithms can tailor the consent content based on the patient's medical history, comprehension level, and specific concerns, thereby enhancing understanding and satisfaction. AI can also automate the updating process of consent documents to ensure compliance with the latest medical guidelines and regulations. Moreover, AI-powered analytics can offer insights into common patient queries or areas of misunderstanding, guiding continuous improvement of the consent process (Di Battista et al., 2023).

The exploration of AI in drafting informed consent documents remains scant (Abou-Abdallah et al., 2024; Kasapovic et al., 2024; Kienzle et al., 2024; Patil et al., 2024; Shiraishi et al., 2024; Szczesniewski et al., 2024) and to date there has been no published research to date on AI's ability to generate informed consent documents for oral and maxillo-facial surgery procedures.

This study aims to evaluate whether standard AI-generated consent forms, provide a viable alternative to traditional, human-generated documents, in terms of accuracy, completeness and readability. These consents were generated by the two most commonly used AI platforms and by a first-year oral surgery resident.

2. Materials and methods

2.1. Working group

In February 2023, an international research group was established, comprising maxillofacial surgeons, dentists, and otolaryngologists from the young sections of the Italian Society of Maxillofacial Surgery and the International Federation of Otorhinolaryngology Societies. The purpose of this consortium was to explore the potential applications, reliability, and possible risks associated with artificial intelligence (AI) platforms in the field of head and neck surgery.

The present study involved 26 researchers from 19 centers across 3 countries. Conducted in alignment with the Helsinki Declaration, the study did not require ethical committee approval as it did not involve patients or animals.

2.2. Informed consent collection

To decide which consents to include in our study, we requested three major oral surgery centers, each performing over 5000 surgical procedures annually, to provide a list of their most frequently conducted surgeries. From this data, we identified the top 10 procedures for inclusion: lower third molar surgical extraction, simple tooth extraction, dental implant placement, apicoectomy, guided bone regeneration, lingual frenulectomy, maxillary sinus lift, full-arch upper jaw rehabilitation with implants, correction of gingival recession using connective tissue graft from the palate, and tooth extraction in patients undergoing anti-resorptive drug therapy.

The framework and contents of the informed consents were determined by a panel of experts, including an oral surgeon, a medical legal expert, and a jurist, based on guidelines for informed consent quality (Kinnersley et al., 2013; Guarino et al., 2022). This framework encompasses the purpose of the procedure, a description of the procedure, expected benefits, possible alternatives, risks and potential complications, and the recovery process.

For the purposes of the study, the two most utilized large language models (LLMs) were employed: ChatGPT4 (OpenAI, San Francisco, CA, US) (https://chat.openai.com) and Google Bard Gemini advanced version (Google Inc., Mountain View, CA, US) (https://gemini.google.com/app). For both LLMs, the same prompt was used, crafted by a panel of experts that included two oral surgeons and an artificial intelligence expert: "Develop an informed consent form for the procedure [name of the procedure], organizing it into the following sections: purpose of the procedure, description of the intervention, expected benefits, possible alternatives, risks and potential complications, recovery process. The consent must be accessible to a patient at a 12th grade reading level." The same researcher entered all the prompts into the chatbots on February 21, 2023. Each prompt was imputted using incognito mode, always in a new chat window.

Additionally, the same prompts were provided to a dentist in his first year of oral surgery residency.

The responses from the LLMs and the resident were collected by the same researcher, randomized, and then sent to the review panel for evaluation [Supplementary Table 1].

2.3. Evaluation protocol

The pool of reviewers consisted of 18 oral and maxillofacial surgeons, each with at least 10 years of experience. Each reviewer independently evaluated each consent form. For each procedure, a randomized assessment of the three consents (the two from the LLMs and the one from the resident) was conducted, evaluating each section of the consent and finally providing an overall judgment. The reviewers assessed the completeness, accuracy, and readability, in addition to

giving an overall judgment. As previously proposed by Decker et al. (2023), the overall judgment for each component of the consent was based on a scale from 0 to 3, depending on whether the component was complete (3), incomplete (2), absent (1), or incorrect (0). Completeness, accuracy, and readability were evaluated using a Likert scale from 0 to 5 [Appendix A].

The readability of each consent form was also assessed by an independent and blinded reviewer using the Gunning Fog Index, a tool that measures the complexity of English text based on sentence length and the number of complex words (Nangia et al., 2022).

2.4. Statistical analysis

Statistical analysis was accomplished using the freely available and open-source software Jamovi, version 2.3.18.0, accessible online at www.jamovi.org. The evaluations of accuracy, completeness, readability, and the overall assessment by the 18 reviewers were reported as median and interquartile range (IQR). Statistical differences between the evaluations of the three consent forms were studied using the Friedman test for paired data. The Durbin-Conover test was used for the post-hoc evaluation of pairwise differences between the consents. Finally, the percentage of consents that were actually written at a 12th-grade reading level or lower was calculated. Differences between the three types of authors were evaluated using Fisher's exact test and Friedman test for paired data. The threshold for statistical significance was established at $p < 0.05$, with a confidence interval of 95%.

3. Results

3.1. Accuracy evaluation results

Human-written consents had a median accuracy score of 4 [IQR 3–4]. ChatGPT-4's consents had a median score of 4 [IQR 4-4], while Gemini Advanced's consents had a median score of 3 [IQR 3–4].

ChatGPT-4 consents were generally more accurate than those produced by both Bard Gemini Advanced ($p < 0.001$) and humans ($p < 0.001$). There were no significant differences in accuracy between Bard and human-written consents ($p = 0.817$) [Fig. 1].

When analyzing individual consents, ChatGPT-4 consistently achieved higher accuracy scores than humans in 5 out of 10 cases and surpassed Bard in 7 out of 10 cases. Bard produced a more accurate consent than humans in one instance, while humans outperformed Bard in another case. Human-written consents were never more accurate than those generated by ChatGPT-4 [Supplementary Table 2].

3.2. Completeness evaluation results

The median completeness score for human-written consents was 4 [IQR 3–4]. ChatGPT-4 generated consents with a median completeness score of 4 [IQR 4–5], while Bard Gemini Advanced produced consents with a median score of 3 [IQR 3–4]. Overall, the consents generated by ChatGPT-4 were significantly more complete compared to those produced by both Bard Gemini Advanced ($p < 0.001$) and the human ($p < 0.001$). There were no statistically significant differences between the completeness scores of Bard Gemini and those of the resident ($p = 0.915$) [Fig. 2].

In the analysis of individual consents, ChatGPT-4 consistently scored significantly higher in completeness than human in 5 out of 10 cases and outperformed Bard in 6 out of 10 cases. The resident provided a significantly more complete consent than Bard Gemini in one case. However, human-written consents were never significantly more complete than those produced by ChatGPT-4 [Supplementary Table 3].

3.3. Readability evaluation results

The median readability score for human-written consents was 4 [IQR 3–4]. ChatGPT-4 generated consents with a median readability score of 4 [IQR 4–5], while Gemini Advanced produced consents with a median score of 4 [IQR 4-4].

Overall, the consents generated by ChatGPT-4 were significantly more readable compared to those produced by both Bard Gemini Advanced ($p < 0.001$) and human ($p < 0.001$). Additionally, Bard's consents were significantly more readable than those written by the resident ($p < 0.001$) [Fig. 3].

In the analysis of individual consents, ChatGPT-4 consistently scored significantly higher in readability than the resident in 8 out of 10 cases and outperformed Bard in 4 out of 10 cases. Bard's consents were significantly more readable than those written by the resident in 4 cases, while human-written consents achieved a significantly better readability score than Bard in only one case. Human-written consents were never significantly more readable than those produced by ChatGPT-4 [Supplementary Table 4].

The median Gunning Fog Index for human-written consents was 20 [IQR 19.2–20.9]. In comparison, consents generated by ChatGPT-4 had a median Gunning Fog Index of 17.2 [IQR 16.5–18.2], while those produced by Bard Gemini Advanced had a median of 23.1 [IQR 20.5–24.7]. None of the consents, whether written by humans or AI chatbots, met the prescribed reading level of 12 (Fisher exact test, $p = 1$). Overall, the consents generated by ChatGPT-4 demonstrated a significantly lower (and therefore closer to the target 12) Gunning Fog Index compared to both the human-written consents ($p = 0.007$) and those by Bard Gemini Advanced ($p < 0.001$). No significant differences

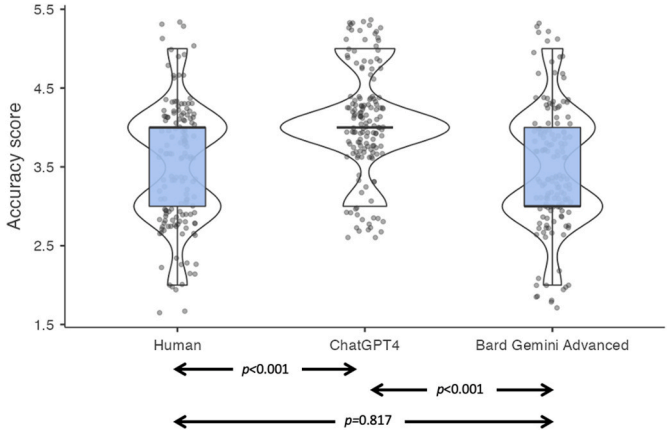

Fig. 1. Accuracy evaluation results.

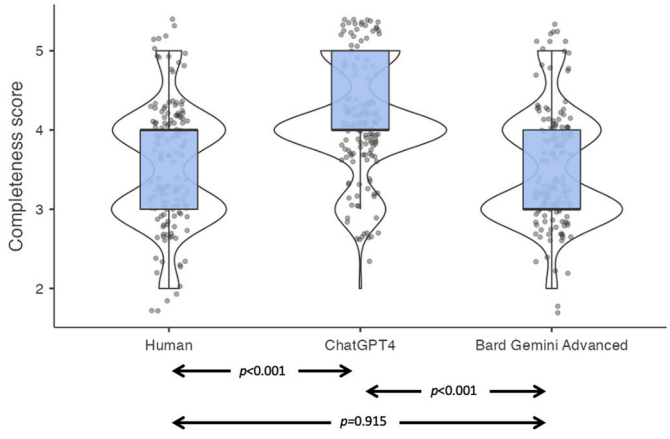

Fig. 2. Completeness evaluation results.

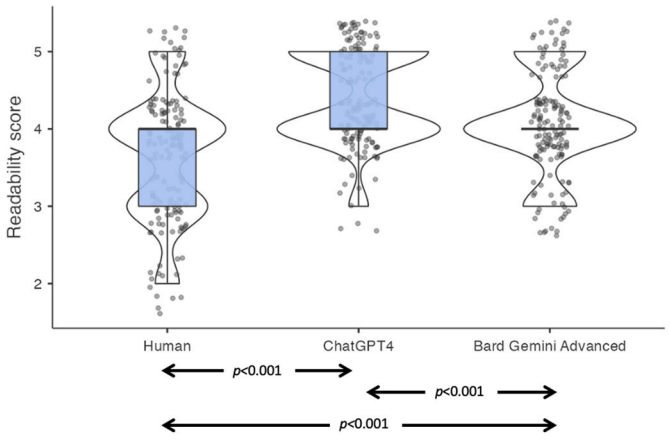

Fig. 3. Readability evaluation results.

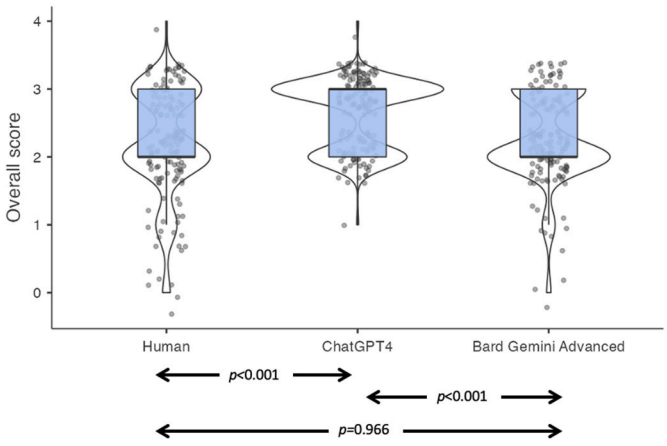

Fig. 5. Overall evaluation results.

were observed between human and Bard Gemini Advanced consents (p = 0.236) [Fig. 4].

3.4. Overall evaluation results

The overall evaluation of the informed consents revealed significant differences in quality among those produced by humans and AI chatbots. The median overall score for human-written consents was 2 [IQR 2–3]. ChatGPT-4 generated consents with a median overall score of 3 [IQR 2–3], while Bard Gemini Advanced produced consents with a median score of 2 [IQR 2–3].

Overall, consents generated by ChatGPT-4 were significantly better compared to those produced by both Bard Gemini Advanced ($p < 0.001$) and resident ($p < 0.001$). There were no statistically significant differences between the consents generated by Bard and those written by the resident ($p = 0.966$) [Fig. 5].

In the analysis of individual consents, ChatGPT-4 consistently scored significantly higher than humans in 4 out of 10 cases and outperformed Bard in 5 out of 10 cases. No significant differences were found between the scores of Bard Gemini Advanced and those of human-written consent in any instance. Additionally, human-generated consents never scored significantly better than those produced by the AI chatbots [Supplementary Table 5].

4. Discussion

The quality of informed consents generated by AI is a relatively unexplored field, with no published research specifically addressing oral surgery. However, several studies have investigated AI-generated information in other medical fields.

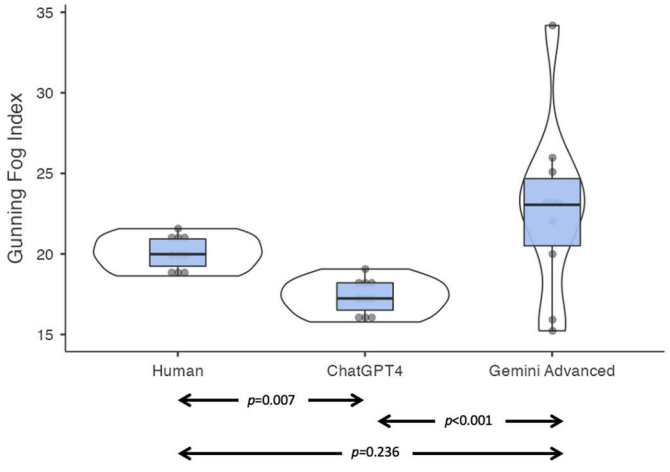

Fig. 4. Gunning fog index results.

In orthopedics and trauma surgery, ChatGPT provided moderate-quality information but showed significant variability in usefulness (Kasapovic et al., 2024). For ENT surgeries, simplified ChatGPT information was more readable but less comprehensive than professional leaflets (Abou-Abdallah et al., 2024). In urology, ChatGPT, Bard, and Copilot offered reasonably accurate information but often lacked comprehensive details and citations (Szczesniewski et al., 2024). In total knee arthroplasty, ChatGPT conveyed accurate information but sometimes generated non-existent references (Kienzle et al., 2024). Ophthalmology research found ChatGPT superior to Bard for preoperative information, yet both failed to fully address adverse events (Patil et al., 2024). For blepharoplasty, AI-generated informed consents were not yet suitable as standalone resources but had potential for improvement (Shiraishi et al., 2024).

The results of the present study indicate that AI-generated consents, particularly those produced by ChatGPT-4, consistently outperformed human-written documents in terms of accuracy, completeness, readability, and overall quality. Our findings revealed that ChatGPT-4-generated consents were generally more accurate than both human-written and Bard-generated consents. The superior performance of ChatGPT-4 can be attributed to its extensive training on a vast dataset, which includes medical literature, allowing it to provide detailed and precise information. This highlights the potential of AI in reducing the risk of misinformation, which is crucial in the context of informed consent where accurate and clear information is paramount for patient decision-making.

The completeness of informed consent documents is essential for ensuring patients have a comprehensive understanding of their procedures, including risks, benefits, and alternatives. ChatGPT-4's consents were more complete compared to both human and Bard-generated consents. This suggests that AI can help overcome the common issue of omitting critical information, which can sometimes occur in human-written consents due to oversight or lack of time (Blackwood et al., 2015). By ensuring all necessary components are included, AI can enhance the integrity of the consent process.

Readability is a crucial factor in ensuring that patients can understand the information provided (Padiya, 2010; Yildiz et al., 2018). While none of the consent documents met the prescribed 12th-grade reading level, ChatGPT-4's consents had a significantly lower Gunning Fog Index compared to both human-written and Bard-generated consents. This indicates that AI can produce more comprehensible documents, potentially improving patient understanding and satisfaction. However, the failure to meet the target reading level across all consents suggests a need for further refinement of AI algorithms to enhance accessibility. The observed discrepancy between the Gunning Fog Index scores and

reviewer ratings can be explained by the fact that these two methods evaluate different aspects of readability. The Gunning Fog Index quantifies linguistic complexity based on sentence length and word difficulty, providing an objective measure (Nangia et al., 2022). However, it does not capture qualitative elements such as the clarity of explanations, logical flow, and contextual appropriateness of medical terminology. Reviewer ratings, on the other hand, reflect human perceptions of readability, considering these qualitative factors (Pandiya, 2010). This suggests that while the Gunning Fog Index offers valuable insights into text complexity, it may not fully align with human assessments of how easily a document can be understood. Consequently, reviewer ratings might identify subtleties in text comprehension that are not apparent through numerical indices alone, emphasizing the importance of combining both objective and subjective measures when evaluating readability.

The overall quality assessment revealed that ChatGPT-4's consents were significantly better than those produced by both Bard and the resident. This aligns with previous studies indicating that advanced AI models can outperform human efforts in specific tasks involving the synthesis and presentation of complex information (Ayers et al., 2023; Hopkins et al., 2023; Lechien et al., 2024; Lorenzi et al., 2024; Vaira et al., 2024a). The consistent performance of ChatGPT-4 underscores its potential as a valuable tool in the creation of informed consent documents, potentially improving the standardization and reliability of patient information.

The differences between the performances of ChatGPT-4 and Bard could be attributed to the variations in their training data and underlying algorithms. ChatGPT-4, benefits from extensive training on a diverse and comprehensive dataset, including a wide array of medical texts. This broad training enables it to generate more detailed and accurate content. In contrast, Bard may have limitations in its training data scope or algorithmic nuances that affect its ability to generate equally detailed and accurate consent documents.

Moreover, AI platforms can provide a more general and objective perspective, less influenced by individual biases or outdated practices that human writers might adhere to. Human-generated consents might reflect the specific training and personal experiences of the healthcare provider, which can sometimes diverge from current evidence-based practices (Blease et al., 2016; Lamont et al., 2019). AI, on the other hand, synthesizes information from a vast array of sources, potentially leading to more standardized and up-to-date content.

The adoption of AI in generating informed consent documents presents several practical advantages. AI can streamline the consent process, reduce the workload on healthcare professionals, and ensure that patients receive up-to-date and comprehensive information. Additionally, AI-powered tools can personalize consent documents to align with individual patient needs, enhancing understanding and engagement. However, the integration of AI in clinical practice should be approached with caution. Continuous evaluation and oversight are necessary to ensure the accuracy and appropriateness of AI-generated content, and healthcare providers must remain involved in the final approval of informed consent documents.

This study has several limitations. First, the evaluation of consents was conducted by a panel of experts, which may introduce subjective bias. No direct feedback was obtained from patients, who are the primary end-users of informed consent documents. This was a deliberate choice, as the current study represents the first phase of a larger research project aimed at evaluating the feasibility and quality of AI-generated consent documents. Initially, we sought to establish a foundational assessment by experienced clinicians to ensure that the content produced was clinically accurate, complete, and appropriate. The inclusion of patient perspectives is planned for the next phase, where we will focus on evaluating patient understanding, satisfaction, and engagement with these documents. Second, the study focused on ten specific procedures, and the generalizability of the findings to other surgical contexts remains to be explored. Third, another limitation of this study is the use of a standardized, identical prompt for all AI-generated consent documents, which does not account for individualized adjustments based on patient-specific factors such as age, literacy level, or clinical history. This approach was chosen to ensure consistency and establish a baseline for evaluating the basic capabilities of AI models in generating informed consent documents. By using a standardized prompt, we were able to conduct a controlled comparison with human-written consents to determine whether AI could produce content of comparable quality without customization. However, we acknowledge that future research should explore the use of more personalized prompts that allow for tailored content. This would enable the generation of individualized consent documents that better address patient-specific needs and concerns, ultimately enhancing patient understanding and engagement. Fourth, another limitation is the absence of comparison with standardized patient information sheets provided by established publishers. These documents typically undergo multiple rounds of expert review and are curated by professionals specialized in medical communication, making them a gold standard for consent documentation. While our study focused on comparing AI-generated texts with those drafted by a first-year resident to establish a performance baseline, future research should include comparisons with standardized documents to fully assess the potential of AI-generated content. Such comparisons would allow for a more robust evaluation of whether AI-generated documents can achieve the same level of detail, accuracy, and clarity as those produced through conventional professional processes.

5. Conclusions

The findings of this study demonstrate that AI platforms, particularly ChatGPT-4, can generate informed consent documents that are more accurate, complete, and readable than those written by a human resident. This suggests a promising role for AI in enhancing patient communication and the informed consent process in healthcare. As AI technology continues to evolve, its integration into clinical practice should be guided by rigorous evaluation and ethical considerations to maximize benefits while safeguarding patient autonomy and understanding.

Future research should focus on refining AI algorithms to further improve readability and exploring the integration of AI-generated consents in diverse clinical settings. Additionally, investigating patient perspectives on AI-generated consent documents can provide valuable insights into their acceptance and effectiveness in enhancing patient understanding and engagement.

Funding

This research was funded by Fondo di Ateneo per la ricerca 2019–2020 of the University of Sassari.

Ethical approval

Not required.

Declaration of competing interest

All authors report no conflict of interest.

Acknowledgement

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcms.2024.10.002.

References

Abou-Abdallah, M., Dar, T., Mahmudzade, Y., Michaels, J., Talwar, R., Tornari, C., 2024. The quality and readability of patient information provided by ChatGPT: can AI reliably explain common ENT operations? Eur. Arch. Oto-Rhino-Laryngol. https://doi.org/10.1007/s00405-024-08598-w.

Ayers, J.W., Poliak, A., Dredze, M., Leas, E.C., Zhu, Z., Kelley, J.B., et al., 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern. Med. 183, 589–596.

Blackwood, R., Maio, R., Mrdjenovich, A., Vandenbosch, T., Gordon, P., Shipman, E., Hamilton, T., 2015. Analysis of the nature of IRB contingencies required for informed consent document approval. Account. Res. 22, 237–245.

Blease, C., Lilienfeld, S., Kelley, J., 2016. Evidence-based practice and psychological treatments: the imperatives of informed consent. Front. Psychol. 10 (7), 1170.

Chiesa-Estomba, C.M., Lechien, J.R., Vaira, L.A., Brunet, A., Cammaroto, G., Mayo-Yanez, M., et al., 2024. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. Eur. Arch. Oto-Rhino-Laryngol. 281, 2081–2086.

Cocanour, C.S., 2017. Informed consent-It's more than a signature on a piece of paper. Am. J. Surg. 214, 993–997.

Decker, H., Trang, K., Ramirez, J., Colley, A., Pierce, L., Coleman, M., et al., 2023. Large Language model-based chatbot vs surgeon-generated informed consent cocumentation for common procedures. JAMA Netw. Open 6, e2336997.

Di Battista, M., Kernitsky, J., Dibart, S., 2023. Artificial intelligence chatbots in patient communication: current possibilities. Int. J. Periodontics Restor. Dent. https://doi.org/10.11607/prd.6925.

Glaser, J., Nouri, S., Fernandez, A., Sudore, R.L., Schillinger, D., Klein-Fedyshin, M., Schenker, Y., 2020. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: an updated systematic review. Med. Decis. Making 40, 119–143.

Guarino, J., Parvanova, I., Finkelstein, J., 2022. Characteristics of electronic informed consent platforms for consenting patients to research studies: a scoping review. Stud. Health Technol. Inf. 290, 777–781.

Hopkins, A.M., Logan, J.M., Kichenadasse, G., Sorich, M.J., 2023. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. JNCI Cancer Spectr. 7 (2), pkad010.

Kasapovic, A., Ali, T., Babasiz, M., Bojko, J., Gathen, M., Kaczmarczyk, R., Roos, J., 2024. Does the information quality of chatGPT meet the requirements of orthopedics and trauma surgery? Cureus 16, e60318.

Kienzle, A., Niemann, M., Meller, S., Gwinner, C., 2024. ChatGPT may offer an adequate substitute for informed consent to patients prior to total knee arthroplasty-Yet caution is needed. J. Personalized Med. 14, 69.

Kinnersley, P., Phillips, K., Savage, K., Kelly, M.J., Farrell, E., Morgan, B., et al., 2013. Interventions to promote informed consent for patients undergoing surgical and other invasive healthcare procedures. Cochrane Database Syst. Rev. 6 (7), CD009445.

Lamont, S., Stewart, C., Chiarella, M., 2019. Capacity and consent: knowledge and practice of legal and healthcare standards. Nurs. Ethics 26, 71–83.

Lechien, J.R., Naunheim, M.R., Maniaci, A., Radulesco, T., Saibene, A.M., Chiesa-Estomba, C.M., Vaira, L.A., 2024. Performance and consistency of ChatGPT-4 versus otolaryngologists: a clinical case series. Otolaryngol. Head Neck Surg. 170, 1519–1526.

Liu, J., Wang, C., Liu, S., 2023. Utility of ChatGPT in clinical practice. J. Med. Internet Res. 25, e48568.

Lorenzi, A., Pugliese, G., Maniaci, A., Lechien, J.R., Allevi, F., Boscolo-Rizzo, P., et al., 2024. Reliability of large language models for advanced head and neck malignancies management: a comparison between ChatGPT 4 and Gemini Advanced. Eur. Arch. Oto-Rhino-Laryngol. https://doi.org/10.1007/s00405-024-08746-2.

Nangia, D., Saini, A., Krishnan, A., Sharma, S., Kumar, V., Chawla, A., Logani, A., 2022. Quality and accuracy of patient-oriented Web-based information regarding tooth avulsion. Dent. Traumatol. 38, 299–308.

Pandiya, A., 2010. Readability and comprehensibility of informed consent forms for clinical trials. Perspect Clin Res 1, 98–100.

Patil, N.S., Huang, R., Mihalache, A., Kisilevsky, E., Kwok, J., Popovic, M.M., et al., 2024. The ability of artificial intelligence chatobots ChatGPT and Google Bard to accurately convey preoperative information for patients undergoing ophthalmic surgeries. Retina 44, 950–953.

Radulesco, T., Saibene, A.M., Michel, J., Vaira, L.A., Lechien, J.R., 2024. ChatGPT-4 performance in rhinology: a clinical case series. Int Forum Allergy Rhinol 14, 1123–1130.

Shiraishi, M., Tomioka, Y., Miyakuni, A., Moriwaki, Y., Yang, R., Oba, J., Okazaki, M., 2024. Generating informed consent documents related to blepharoplasty using ChatGPT. Ophthalmic Plast. Reconstr. Surg. 40, 316–320.

Szczesniewski, J.J., Ramos Alba, A., Rodríguez Castro, P.M., Lorenzo Gómez, M.F., Sainz González, J., Llanes González, L., 2024. Quality of information about urologic pathology in English and Spanish from ChatGPT, BARD, and Copilot. Actas Urol. Esp. 48, 398–403.

Vaira, L.A., Lechien, J.R., Abbate, V., Allevi, F., Audino, G., Beltramini, G.A., et al., 2024. Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: a multicenter collaborative analysis. Otolaryngol. Head Neck Surg. 170, 1492–1503.

Vaira, L.A., Lechien, J.R., Abbate, V., Allevi, F., Audino, G., Beltramini, G.A., et al., 2024a. Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool: a new tool to assess the quality of health information provided by AI platforms. Eur. Arch. Oto-Rhino-Laryngol. https://doi.org/10.1007/s00405-024-08710-0.

Yildiz, M., Kozanhan, B., Tutar, M., 2018. Assessment of readability level of informed consent forms used in intensive care units. Med. Sci. 8, 277–281.