ChatGPT for Neurosurgical Patient Education

- Assessing AI Simplification of Medical Texts: Readability and Content Fidelity
- Artificial intelligence as a modality to enhance the readability of neurosurgical literature for patients
- Text-to-video generative artificial intelligence: sora in neurosurgery
- Usefulness and Accuracy of Artificial Intelligence Chatbot Responses to Patient Questions for Neurosurgical Procedures

The findings demonstrate promising potential for the application of the ChatGPT in patient education. GPT4 is an accessible tool that can be an immediate solution to enhancing the readability of current neurosurgical literature. Layperson summaries generated by GPT4 would be a valuable addition to neurosurgical journals and would be likely to improve comprehension for patients using internet resources like PubMed¹⁾.

The escalating complexity of medical literature necessitates tools to enhance readability for patients. This study aimed to evaluate the efficacy of ChatGPT-4 in simplifying neurology and neurosurgical abstracts and patient education materials (PEMs) while assessing content preservation using Latent Semantic Analysis (LSA).

Methods: A total of 100 abstracts (25 each from Neurosurgery, Journal of Neurosurgery, Lancet Neurology, and JAMA Neurology) and 340 PEMs (66 from the American Association of Neurological Surgeons, 274 from the American Academy of Neurology) were transformed by a GPT-4.0 prompt requesting a 5th grade reading level. Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease (FKRE) scores were used before/after transformation. Content fidelity was validated via LSA (ranging 0-1, 1 meaning identical topics) and by expert assessment (0-1) for a subset (n = 40). Pearson correlation coefficient compared assessments.

Results: FKGL decreased from 12th to 5th grade for abstracts and 13th to 5th for PEMs (p < 0.001). FKRE scores showed similar improvement (p < 0.001). LSA confirmed high content similarity for abstracts (mean cosine similarity 0.746) and PEMs (mean 0.953). The expert assessment indicated a mean topic similarity of 0.775 for abstracts and 0.715 for PEMs. The Pearson coefficient between LSA and expert assessment of textual similarity was 0.598 for abstracts and -0.167 for PEMs. Segmented analysis of similarity correlations revealed a correlation of 0.48 (p = 0.02) below 450 words and a -0.20 (p = 0.43) correlation above 450 words.

GPT-4.0 markedly improved the readability of medical texts, predominantly maintaining content integrity as substantiated by LSA and expert evaluations. LSA emerged as a reliable tool for assessing content fidelity within moderate-length texts, but its utility diminished for longer documents, overestimating similarity. These findings support the potential of AI in combating low health literacy, however, the similarity scores indicate expert validation is crucial. Future research must strive to improve transformation precision and develop validation methodologies².

Guerra et al. in a study, assessed the ability of Chat Generative Pretrained Transformer (ChatGPT) 3.5 and ChatGPT4 to generate readable and accurate summaries of published neurosurgical literature.

Abstracts published in journal issues released between June 2023 and August 2023 (n = 150) were randomly selected from the top 5 ranked neurosurgical journals according to Google Scholar. ChatGPT models were instructed to generate a readable layperson summary of the original abstract from a statistically validated prompt. Readability results and grade-level indicators (RR-GLIs) scores were calculated for GPT3.5- and GPT4-generated summaries and original abstracts. Two physicians independently rated the accuracy of ChatGPT-generated layperson summaries to assess scientific validity. One-way ANOVA followed by a pairwise t-test with Bonferroni correction was performed to compare readability scores. Cohen's kappa was used to assess interrater agreement between the two rater physicians.

Analysis of 150 original abstracts showed a statistically significant difference for all RR-GLIs between the ChatGPT-generated summaries and original abstracts. The readability scores are formatted as follows (original abstract mean, GPT3.5 summary mean, GPT4 summary mean, p-value): Flesch-Kincaid reading grade (12.55, 7.80, 7.70, p < 0.0001); Gunning fog score (15.46, 10.00, 9.00, p < 0.0001); Simple Measure of Gobbledygook (SMOG) index (11.30, 7.13, 6.60, p < 0.0001); Coleman-Liau index (14.67, 11.32, 10.26, p < 0.0001); automated readability index (10.87, 8.50, 7.75, p < 0.0001); and Flesch-Kincaid reading ease (33.29, 68.45, 69.55, p < 0.0001). GPT4-generated summaries demonstrated higher RR-GLIs than GPT3.5-generated summaries in the following categories: Gunning fog score (0.0003); SMOG index (0.027); Coleman-Liau index (< 0.0001); sentences (< 0.0001); complex words (< 0.0001); and % complex words (0.0035). A total of 68.4% and 84.2% of GPT3.5- and GPT4-generated summaries, respectively, maintained moderate scientific accuracy according to the two physician reviewers.

The findings demonstrate promising potential for the application of the ChatGPT in patient education. GPT4 is an accessible tool that can be an immediate solution to enhancing the readability of current neurosurgical literature. Layperson summaries generated by GPT4 would be a valuable addition to a neurosurgical journal and would be likely to improve comprehension for patients using internet resources like PubMed³⁾.

Systematic reviews

A review aims to examine the accuracy, reliability, comprehensiveness and readability of medical patient education materials (PEMs) simplified by AI models. A systematic review was conducted searching for articles assessing outcomes of use of AI in simplifying patient education materials (PEMs). Inclusion criteria are as follows: publication between January 2019 and June 2023, various modalities of AI, English language, AI use in PEMs and including physicians and/or patients. An inductive thematic approach was utilised to code for unifying topics which were qualitatively analysed. Twenty studies were included, and seven themes were identified (reproducibility, accessibility and ease of use, emotional support and user satisfaction, readability, data security, accuracy and reliability and comprehensiveness). AI effectively simplified PEMs, with reproducibility rates up to 90.7% in specific domains. User satisfaction exceeded 85% in AI-generated materials. AI models showed promising readability improvements, with ChatGPT achieving 100% post-simplification readability scores. AI's performance in accuracy and reliability was mixed, with occasional lack of comprehensiveness and inaccuracies, particularly when addressing complex medical topics. AI models accurately simplified basic tasks but lacked soft skills and personalisation. These limitations

can be addressed with higher-calibre models combined with prompt engineering. In conclusion, the literature reveals a scope for AI to enhance patient health literacy through medical PEMs. Further refinement is needed to improve AI's accuracy and reliability, especially when simplifying complex medical information ⁴⁾.

The review highlights the transformative potential of AI in simplifying PEMs and enhancing patient health literacy. However, limitations in the scope, depth, and evaluation criteria constrain its findings. While AI excels in readability and user satisfaction, challenges like accuracy, comprehensiveness, and ethical considerations demand further research and refinement. Addressing these gaps is essential for realizing the full potential of AI in medical education and patient care.

Quantitative experimental research studies

The escalating complexity of medical literature necessitates tools to enhance readability for patients. This study aimed to evaluate the efficacy of ChatGPT-4 in simplifying neurology and neurosurgical abstracts and patient education materials (PEMs) while assessing content preservation using Latent Semantic Analysis (LSA).

Picton et al. total of 100 abstracts (25 each from Neurosurgery, Journal of Neurosurgery, Lancet Neurology, and JAMA Neurology) and 340 PEMs (66 from the American Association of Neurological Surgeons, 274 from the American Academy of Neurology) were transformed by a GPT-4.0 prompt requesting a 5th grade reading level. Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease (FKRE) scores were used before/after transformation. Content fidelity was validated via LSA (ranging 0-1, 1 meaning identical topics) and by expert assessment (0-1) for a subset (n = 40). Pearson correlation coefficient compared assessments.

FKGL decreased from 12th to 5th grade for abstracts and 13th to 5th for PEMs (p < 0.001). FKRE scores showed similar improvement (p < 0.001). LSA confirmed high content similarity for abstracts (mean cosine similarity 0.746) and PEMs (mean 0.953). Expert assessment indicated a mean topic similarity of 0.775 for abstracts and 0.715 for PEMs. The Pearson coefficient between LSA and expert assessment of textual similarity was 0.598 for abstracts and -0.167 for PEMs. Segmented analysis of similarity correlations revealed a correlation of 0.48 (p = 0.02) below 450 words and a -0.20 (p = 0.43) correlation above 450 words.

GPT-4.0 markedly improved the readability of medical texts, predominantly maintaining content integrity as substantiated by LSA and expert evaluations. LSA emerged as a reliable tool for assessing content fidelity within moderate-length texts, but its utility diminished for longer documents, overestimating similarity. These findings support the potential of AI in combating low health literacy, however, the similarity scores indicate expert validation is crucial. Future research must strive to improve transformation precision and develop validation methodologies⁵⁾

The study provides compelling evidence that GPT-4.0 can enhance the readability of medical literature while largely preserving content. However, it also highlights the limitations of automated tools in fully capturing content fidelity, particularly for longer texts. Future research should:

Explore methods to refine AI transformations for nuanced accuracy. Investigate the impact of

simplified materials on patient comprehension and health outcomes. Develop hybrid workflows that integrate AI simplification with efficient expert review. This research sets a strong foundation for leveraging AI in patient education but underscores the need for continued development and expert oversight to ensure accuracy and trustworthiness.

Best ChatGPT prompts for Neurosurgical Patient Education

Best ChatGPT prompts for Neurosurgical Patient Education

1)

Guerra GA, Grove S, Le J, Hofmann HL, Shah I, Bhagavatula S, Fixman B, Gomez D, Hopkins B, Dallas J, Cacciamani G, Peterson R, Zada G. Artificial intelligence as a modality to enhance the readability of neurosurgical literature for patients. J Neurosurg. 2024 Nov 8:1-7. doi: 10.3171/2024.6.JNS24617. Epub ahead of print. PMID: 39504543.

2) 5)

Picton B, Andalib S, Spina A, Camp B, Solomon SS, Liang J, Chen PM, Chen JW, Hsu FP, Oh MY. Assessing AI Simplification of Medical Texts: Readability and Content Fidelity. Int J Med Inform. 2024 Dec 1;195:105743. doi: 10.1016/j.ijmedinf.2024.105743. Epub ahead of print. PMID: 39667051.

Guerra GA, Grove S, Le J, Hofmann HL, Shah I, Bhagavatula S, Fixman B, Gomez D, Hopkins B, Dallas J, Cacciamani G, Peterson R, Zada G. Artificial intelligence as a modality to enhance the readability of neurosurgical literature for patients. J Neurosurg. 2024 Nov 8:1-7. doi: 10.3171/2024.6.JNS24617. Epub ahead of print. PMID: 39504543.

Nasra M, Jaffri R, Pavlin-Premrl D, Kok HK, Khabaza A, Barras C, Slater LA, Yazdabadi A, Moore J, Russell J, Smith P, Chandra RV, Brooks M, Jhamb A, Chong W, Maingard J, Asadi H. Can artificial intelligence improve patient educational material readability? A systematic review and narrative synthesis. Intern Med J. 2024 Dec 25. doi: 10.1111/imj.16607. Epub ahead of print. PMID: 39720869.

From: https://neurosurgerywiki.com/wiki/ - **Neurosurgery Wiki**

Permanent link: https://neurosurgerywiki.com/wiki/doku.php?id=chatgpt_for_neurosurgical_patient_education



Last update: 2025/01/20 08:20