

Are Artificial Intelligence-generated Patient Leaflets Ready for Clinical Use? A Readability Comparison across Common Orthopaedic Procedures

Jamila Tukur Jido¹, Ahmed Al Wizni²

Learning Point of the Article:

This study highlights that while artificial intelligence platforms such as ChatGPT, Gemini, and DeepSeek can generate patient information leaflets rapidly, their default outputs are written at a level far above recommended health literacy standards. In contrast, professionally authored materials from the Royal College of Surgeons of England consistently meet readability guidelines, making them more suitable for supporting informed consent and patient engagement. Clinicians and educators should be aware that AI outputs require careful review and adaptation before clinical use.

Abstract

Introduction: Readable patient information is central to informed consent, shared decision-making, and treatment adherence. With the emergence of large language models (LLMs), such as ChatGPT, Gemini, and DeepSeek, there is growing interest in their role in generating health education content. However, the readability of such AI-generated patient information leaflets (PILs) has not been systematically compared with that of professionally authored materials.

Objective: This study aimed to compare the readability of PILs generated by three generative artificial intelligence (AI) platforms with those produced by the Royal College of Surgeons of England (RCS England) for three common orthopedic procedures: Carpal tunnel release, total hip replacement, and total knee replacement.

Materials and Methods: A total of 12 PILs (four per procedure) were analyzed using five validated readability metrics: Flesch reading ease, Flesch-Kincaid grade level, Gunning Fog Index, simple measure of Gobbledygook (SMOG) Index, and Coleman-Liau Index. Each AI model was prompted with a standardized instruction to generate a leaflet for the specified procedure. The RCS England leaflets served as the professional benchmark.

Results: Across all metrics and procedures, RCS England leaflets demonstrated superior readability, with Flesch Reading Ease scores above 70 and Flesch-Kincaid Grade Levels between 5.52 and 7.15. In contrast, AI-generated leaflets frequently exceeded recommended complexity thresholds, with Grade Levels often above 12 and Gunning Fog and SMOG scores indicating post-secondary reading requirements. ChatGPT outputs were the most linguistically complex, while Gemini and DeepSeek produced intermediate but still suboptimal readability.

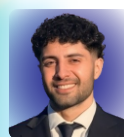
Conclusion: While LLMs offer promising avenues for scalable health communication, current AI-generated PILs do not consistently meet recommended readability standards. Professionally authored leaflets remain more accessible for the average patient. These findings highlight the ongoing need for clinician oversight and quality assurance when integrating AI into patient education materials.

Keywords: Patient information leaflets, readability, artificial intelligence, large language models, health literacy, orthopedic surgery, ChatGPT, Gemini, Deepseek.

Author's Photo Gallery



Dr. Jamila Tukur Jido



Dr. Ahmed Al Wizni

Access this article online

Website:
www.jocr.co.in

DOI:
<https://doi.org/10.13107/jocr.2025.v15.i10.6240>

¹Department of Geriatric Medicine, Queen Elizabeth University Hospital, Glasgow, United Kingdom,
²Department of Surgery, Broomfield Hospital, Broomfield, United Kingdom

Address of Correspondence:

Dr. Jamila Tukur Jido,
Department of Geriatric Medicine, Queen Elizabeth University Hospital, Glasgow, United Kingdom.
E-mail: jamilatukurjido@gmail.com

Submitted: 25/07/2025; Review: 11/08/2025; Accepted: September 2025; Published: October 2025

DOI: <https://doi.org/10.13107/jocr.2025.v15.i10.6240>

© The Author(s). 2025 Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and non-commercial reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

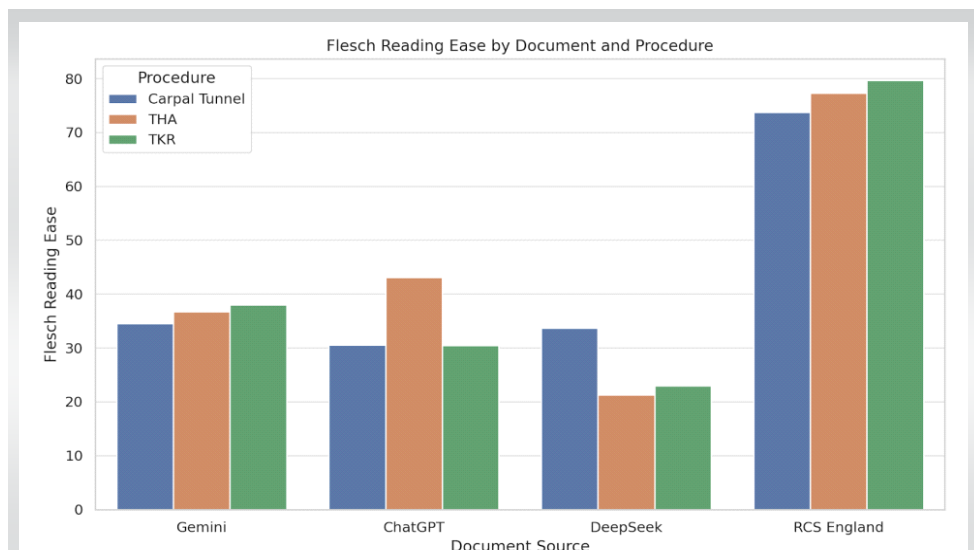


Figure 1: Flesch reading ease by document and procedure. Bar chart comparing Flesch Reading Ease scores for carpal tunnel release, total hip replacement, and total knee replacement leaflets across four sources: ChatGPT, Gemini, DeepSeek, and the Royal College of Surgeons of England (RCS England). Higher scores indicate easier readability.

Introduction

A patient's ability to understand their medical condition, treatment options, and likely outcomes is fundamental to safe, effective, and person-centred care [1, 2]. Comprehension of written health information underpins key aspects of the patient journey, including informed consent, adherence to treatment, shared decision-making, and overall satisfaction [3, 4]. One of the principal tools for communicating such information is the patient information leaflet (PIL), which provides an accessible summary of medical or surgical interventions [5].

Generative artificial intelligence (AI) has emerged as a potentially transformative tool in health communication [6]. Large language models (LLMs) such as ChatGPT (OpenAI), Gemini (Google DeepMind), and DeepSeek have

demonstrated a high degree of fluency in generating human-like text across diverse domains [7]. Early studies indicate that AI-generated content is often both accurate and more readable or empathetic than conventional clinical text [8, 9, 10]. However, there is limited empirical evidence comparing the readability of AI-generated PILs with those developed through established professional processes.

Orthopaedic surgery provides a useful context in which to explore this question. Elective procedures such as carpal tunnel release, total hip replacement (THR), and total knee replacement (TKR) are among the most commonly performed globally and carry substantial implications for pain, mobility, function, and recovery [11]. Patients undergoing these operations frequently access or are given detailed written guidance about perioperative expectations [12]. Ensuring that this material is comprehensible and accessible is essential to optimizing engagement and preparation.

This study aims to address the current evidence gap by evaluating the readability of AI-generated PILs for carpal tunnel release, THR, and TKR. For each procedure, we analyzed four leaflets: One professionally authored by the RCS England and three generated using ChatGPT, Gemini, and DeepSeek. The primary objective was to determine whether significant differences in readability exist between professionally authored and AI-generated texts.

Five validated readability metrics were used to assess textual complexity. The Flesch Reading Ease score evaluates sentence length and syllable count, while the Flesch-Kincaid grade level estimates the U.S. school grade needed for comprehension [13]. The SMOG Index and Gunning Fog Index emphasize polysyllabic word use, correlating with difficulty for readers with lower literacy. The Coleman-Liau Index, designed for automated processing, assesses character count and sentence structure rather than syllables [14].

To the best of our knowledge, this is one of the first studies to systematically evaluate the readability of AI-generated PILs against those produced by a national surgical body. The findings contribute to an evolving discourse on the role of AI in healthcare

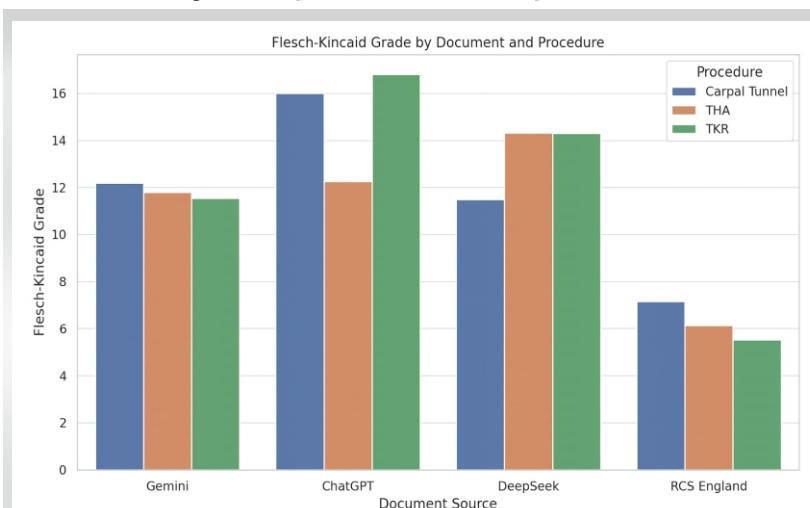


Figure 2: Flesch-Kincaid grade level by document and procedure. Bar chart showing Flesch-Kincaid grade level across procedures and document sources. This metric estimates the U.S. school grade level required to understand the text.

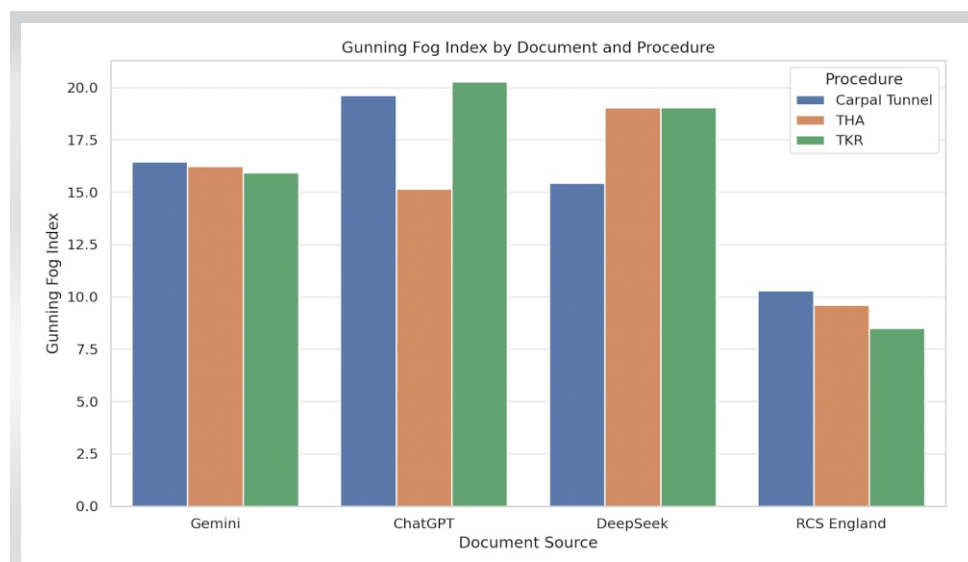


Figure 3: Gunning Fog index by document and procedure. Bar chart comparing the Gunning Fog Index across the four sources for each orthopaedic procedure. Higher scores indicate more years of formal education required to understand the text.

communication, with particular relevance to patient education and health literacy.

Methods

This study was a cross-sectional comparative analysis designed to assess the readability of PILs relating to three common orthopedic procedures: Carpal tunnel release, THR, and TKR. The aim was to compare the accessibility of language between traditionally authored materials and those generated using AI, using established readability formulas. As this study used only publicly available documents and did not involve identifiable human or patient data, ethical approval was not required.

For each of the three procedures, four leaflets were selected.

text output was used for analysis. Any formatting, headings, images, hyperlinks, or metadata were removed to standardize the documents. The resulting clean text was used in all further analyses.

Five different readability formulas were used to assess the leaflets. These included the Flesch Reading Ease score, the Flesch-Kincaid Grade Level, the Gunning Fog Index, the SMOG Index, and the Coleman-Liau Index. These formulas are widely accepted in the field of health communication, and each one assesses different aspects of readability, such as word and sentence length, syllable count, or vocabulary complexity. The readability scores for each leaflet were calculated using the textstat package in Python (version 3.10), which applies the

original algorithms behind each formula.

Once readability scores were generated, basic descriptive statistics were calculated for each leaflet across the three procedure types. The data generated were non-parametric in distribution. However, due to the small sample size ($n = 4$ per procedure), further statistical analysis was not pursued, as it would be unlikely to yield meaningful or generalizable results. Instead, this study serves as a preliminary exploration of the linguistic complexity of AI-generated healthcare materials and

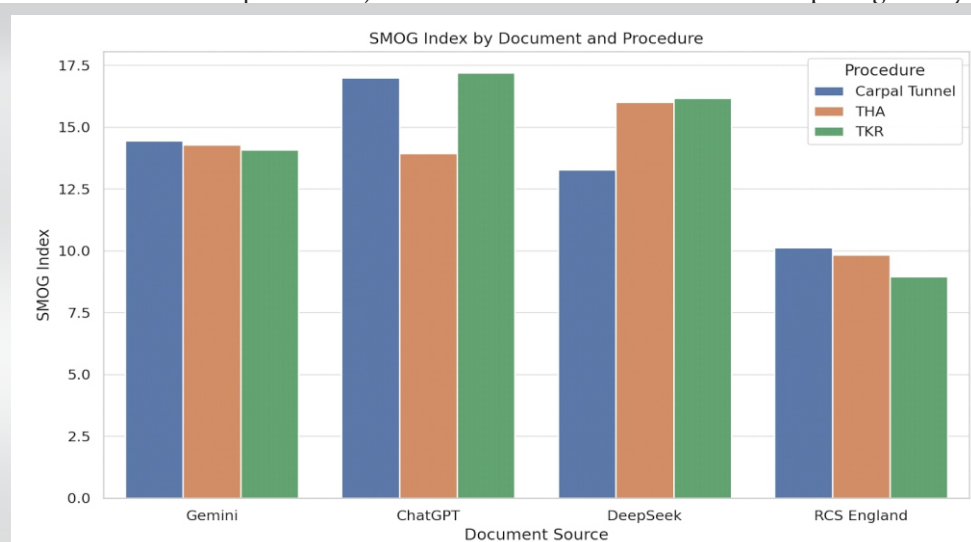


Figure 4: Simple measure of Gobbledygook (SMOG) Index by document and procedure. Bar chart displaying SMOG Index scores for each leaflet and procedure. The SMOG Index estimates the years of education needed to understand a passage based on polysyllabic word count.

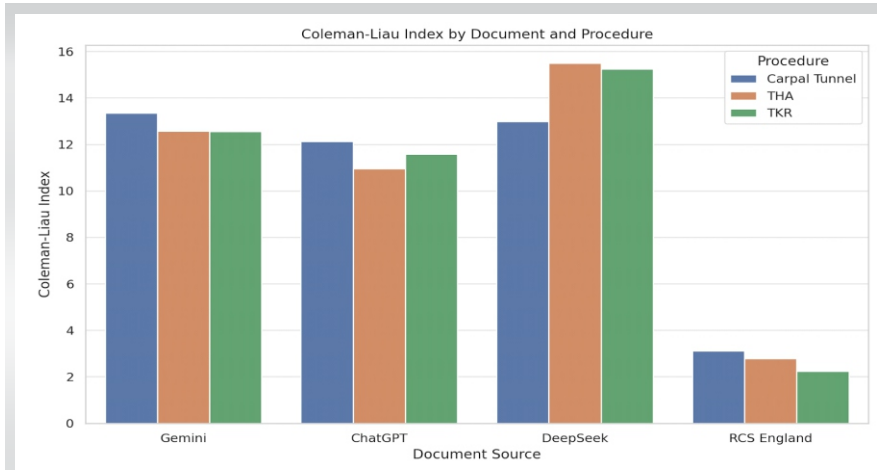


Figure 5: Coleman-Liau Index by document and procedure. Bar chart comparing Coleman-Liau Index scores across document sources for each procedure. This index estimates readability based on characters per word and sentence length.

aims to inform future work on the optimization and safe implementation of AI in patient-facing resources.

Results

Readability scores were calculated for PILs relating to three orthopedic procedures across four sources.

Carpal tunnel release

Readability scores for the carpal tunnel release leaflets are summarized in Table 1. The leaflet from the RCS England was the most readable across all five metrics. It achieved the highest Flesch Reading Ease score (73.71), indicating that it was significantly easier to read than the AI-generated texts. In contrast, the leaflet produced by ChatGPT was the most complex, with the lowest Flesch Reading Ease score (30.50), the highest Flesch-Kincaid Grade Level (15.99), and the

highest SMOG Index (16.99). DeepSeek and Gemini generated intermediate scores, although both exceeded the recommended readability threshold for healthcare materials, which is typically below the U.S. 8th-grade level. The Coleman-Liau Index was also considerably lower for the RCS England leaflet (3.11), compared with the AI-generated leaflets, whose scores ranged from 12.13 to 13.34.

Total hip arthroplasty

As shown in Table 2, similar trends were observed in the THR leaflets. The RCS England leaflet again demonstrated the highest readability, with the highest Flesch Reading Ease score (77.26) and the lowest scores across all other indices: Flesch-Kincaid Grade Level (6.14), Gunning Fog Index (9.61), SMOG Index (9.82), and Coleman-Liau Index (2.79). In contrast, DeepSeek produced the least readable text, with a Flesch Reading Ease score of 21.24 and a Flesch-Kincaid Grade Level of 14.31. It also scored the highest on both the Gunning Fog Index (19.03) and the Coleman-Liau Index (15.50), reflecting its elevated linguistic complexity. ChatGPT and Gemini generated leaflets of intermediate complexity, with Flesch-Kincaid Grade Levels of 12.25 and 11.79, respectively.

TKR

Table 3 presents the readability data for the TKR PILs. As with the other procedures, the RCS England leaflet was the most readable, with a Flesch Reading Ease score of 79.70 and the lowest Flesch-Kincaid Grade Level (5.52). Among the AI-generated texts, ChatGPT produced the least readable content, with a Flesch-Kincaid Grade Level of 16.81 and a Gunning Fog Index of 20.28 – the highest values recorded across all procedure groups. DeepSeek scored highest on the Coleman-Liau Index (15.25) and the SMOG Index (16.17), again indicating elevated linguistic complexity. Gemini's output was more readable than the other AI-generated leaflets but remained more complex than the RCS England document, with a Flesch Reading Ease of 38.01 and a Flesch-Kincaid Grade Level of 11.54.

Comparison of Readability metrics

Fig. 1 illustrates the Flesch Reading Ease scores across all procedures and document sources. In each case,

Table 1: Carpal tunnel release

| Document | Flesch reading ease | Flesch-Kincaid grade | Gunning fog index | SMOG index | Coleman-Liau index |
|-------------|---------------------|----------------------|-------------------|------------|--------------------|
| Gemini | 34.52 | 12.19 | 16.44 | 14.44 | 13.34 |
| ChatGPT | 30.5 | 15.99 | 19.62 | 16.99 | 12.13 |
| DeepSeek | 33.67 | 11.48 | 15.44 | 13.27 | 12.98 |
| RCS England | 73.71 | 7.15 | 10.29 | 10.13 | 3.11 |

SMOG: Simple measure of Gobbledygook, RCS: Royal college of surgeons

Table 2: Total hip arthroplasty

| Document | Flesch reading ease | Flesch-Kincaid grade | Gunning Fog index | SMOG index | Coleman-Liau index |
|--|---------------------|----------------------|-------------------|------------|--------------------|
| Gemini | 36.68 | 11.79 | 16.24 | 14.28 | 12.57 |
| ChatGPT | 43.1 | 12.25 | 15.16 | 13.93 | 10.96 |
| DeepSeek | 21.24 | 14.31 | 19.03 | 16 | 15.5 |
| RCS England | 77.26 | 6.14 | 9.61 | 9.82 | 2.79 |
| SMOG: Simple measure of Gobbledygook, RCS: Royal college of surgeons | | | | | |

the RCS England leaflets had the highest readability scores, exceeding 70 in all three procedures, which indicates text that is easy to read. By contrast, all AI-generated documents scored below 45, falling within the “difficult” category. The greatest discrepancy was observed in the TKR group, where the RCS England leaflet scored 79.7 compared to a low of 22.98 for DeepSeek.

Fig. 2 shows the RCS England documents consistently exhibited the lowest Flesch-Kincaid Grade Levels, ranging from 5.52 to 7.15, indicating a reading age of approximately 11–13 years. In contrast, the AI-generated documents often required comprehension at a college reading level, with ChatGPT producing the most complex text in the TKR group (Grade 16.81). These findings suggest that only the RCS England materials align with health literacy recommendations.

Fig. 3 presents the Gunning Fog Index, with the RCS England leaflets again demonstrating superior readability (scores ranging from 8.49 to 10.29). AI-generated documents consistently exceeded a Fog Index of 15, indicating that they may not be easily understood without post-secondary education. ChatGPT’s leaflet for TKR had the highest score (20.28), suggesting it would be especially difficult for the general population to comprehend.

As shown in Fig. 4, the SMOG Index results mirrored those of the other metrics. RCS England leaflets consistently scored below 11, while AI-generated texts scored between 13 and 17. Notably, the ChatGPT leaflet for carpal tunnel release scored the highest overall (16.99), reinforcing concerns about the density and complexity of its language for the average patient.

Fig. 5 depicts the Coleman-Liau Index across the different procedures and document types. The RCS England leaflets had scores between 2.23 and 3.11, consistent with early secondary school readability. In contrast, AI-generated texts scored significantly higher, with DeepSeek reaching a Coleman-Liau

Table 3: Total knee replacement

| Document | Flesch reading ease | Flesch-Kincaid grade | Gunning Fog index | SMOG index | Coleman-Liau index |
|--|---------------------|----------------------|-------------------|------------|--------------------|
| Gemini | 38.01 | 11.54 | 15.94 | 14.07 | 12.56 |
| ChatGPT | 30.39 | 16.81 | 20.28 | 17.2 | 11.58 |
| DeepSeek | 22.98 | 14.3 | 19.03 | 16.17 | 15.25 |
| RCS England | 79.7 | 5.52 | 8.49 | 8.95 | 2.23 |
| SMOG: Simple measure of Gobbledygook, RCS: Royal college of surgeons | | | | | |

Index of 15.5 for THR, indicative of college-level reading. These results reaffirm that traditional, professionally authored leaflets are more accessible in terms of sentence and word structure.

Across all three procedures, the RCS England leaflets consistently demonstrated the highest readability and the lowest linguistic complexity across all metrics. These documents achieved Flesch Reading Ease scores ranging from 73.71 to 79.70, substantially higher than those of the AI-generated texts, which typically scored below 45. In the TKR group, for example, the RCS England leaflet scored 79.70 compared to just 22.98 for DeepSeek.

Flesch-Kincaid Grade level scores revealed a similar trend. RCS England documents were written at a grade level between 5.52 and 7.15, whereas AI-generated texts frequently exceeded grade 11, with ChatGPT reaching 16.81 in the TKR group. Gunning Fog Index and SMOG Index scores further reinforced this pattern: AI-generated content consistently demonstrated higher levels of word complexity, particularly polysyllabic usage, than the professionally authored documents. The RCS England leaflets maintained Fog and SMOG values around 9 to 10, while AI-generated documents ranged from 13 to over 17.

The Coleman-Liau Index, which is based on characters per word and sentence length, also favored the RCS England leaflets, with scores ranging from 2.23 to 3.11. In contrast, AI-generated documents often scored between 11 and 15, again suggesting a significantly more complex reading level.

Discussion

This study compared the readability of PILs relating to three commonly performed orthopedic procedures authored either by LLMs or the RCS England. Across all three procedures and five validated readability metrics, the RCS England leaflets consistently demonstrated superior readability, achieving

significantly lower complexity scores that align with recommended standards for patient-facing health communication. While the small sample size limited the power of statistical analysis, descriptive trends were uniform and clinically meaningful.

These findings offer preliminary evidence that current AI-generated patient information may not yet meet accepted standards of readability. The professionally authored leaflets from RCS England achieved Flesch-Kincaid Grade Levels ranging from 5.52 to 7.15 and SMOG Index scores below 11, consistent with the comprehension level expected of individuals aged 11–13 years. By contrast, AI-generated leaflets routinely exceeded grade level 13 and recorded high Gunning Fog and SMOG scores, suggesting their suitability only for readers with post-secondary education. Gemini produced moderately more readable outputs, but still fell short of the benchmark set by RCS England.

This discrepancy is significant in the context of health literacy. NHS England and international bodies such as the World Health Organization recommend that patient education materials be pitched at a level accessible to individuals with reading skills equivalent to a U.S. 6th to 8th grade level [15, 16, 17]. However, a substantial proportion of the UK population does not meet this threshold. The 2011 Skills for Life Survey estimated that up to 43% of working-age adults in the UK lack the literacy skills necessary to fully understand written health information [18]. Such disparities in reading ability present a major barrier to equitable healthcare, as patients may struggle to understand their diagnoses, treatment options, or post-operative instructions [19].

The emergence of generative AI platforms such as ChatGPT, Gemini, and DeepSeek offers an exciting opportunity to scale the production of customized health content. These tools are already being used informally by patients to generate explanations of diagnoses, medications, and surgical interventions [20]. However, our study suggests that their default outputs may not yet be suitable for direct patient use without post-processing.

It is important to recognize that the strength of professionally authored materials, such as those from RCS England, may lie in their development process. These leaflets typically benefit from multidisciplinary authorship, peer review, clinical endorsement, and adherence to national guidelines. Their superior readability may reflect deliberate linguistic and structural choices that AI models, at present, are not consistently capable of replicating [8].

This study has several strengths. It is among the first to directly compare multiple AI platforms against a national professional standard across several surgical conditions. The use of five

established readability metrics allows for a nuanced assessment of language complexity. In addition, the focus on three high-volume orthopedic procedures enhances the real-world relevance of the findings, as these conditions frequently require clear patient education in both primary and secondary care settings.

Nonetheless, several limitations must be acknowledged. The sample size was small ($n = 4$ per procedure), limiting statistical power. Furthermore, readability metrics focus primarily on surface-level linguistic features such as sentence length and syllable count. They do not capture other important aspects of communication, such as tone, layout, cultural relevance, or factual accuracy. In addition, the prompts used to generate AI content were intentionally simple to reflect real-world use, but may not represent the optimal input for generating high-quality, readable texts. It is also important to note that readability scores evaluate textual complexity rather than actual comprehension. While they offer insight into accessibility, they do not assess how well patients understand or retain the material. Further research could investigate whether prompt engineering, post-editing by clinicians, or future iterations of LLMs might produce more accessible outputs.

Finally, although the study focused on orthopedic surgical procedures, the findings may have broader implications. Similar evaluations in other clinical areas could further illuminate the potential and limitations of AI in health communication. Qualitative research could also explore patient perceptions, understanding, and trust in AI-generated material.

Conclusion

This study demonstrates that professionally authored PILs from the RCS England outperform AI-generated alternatives in terms of readability. While generative AI holds significant promise for the future of personalized health communication, there remains a critical need for clinician oversight, expert editing, and national guidance to ensure outputs are accessible to all patients. Until AI tools can reliably meet health literacy standards, professionally written and peer-reviewed materials remain essential for safe, effective, and equitable patient education.

Clinical Message

AI-generated patient information leaflets are not yet ready to replace professionally authored materials in routine clinical practice. Until readability and accessibility are improved, clinician oversight remains essential to ensure patients receive information that is clear, comprehensible, and safe for decision-making.

Declaration of patient consent: The authors certify that they have obtained all appropriate patient consent forms. In the form, the patient has given the consent for his/ her images and other clinical information to be reported in the journal. The patient understands that his/ her names and initials will not be published and due efforts will be made to conceal their identity, but anonymity cannot be guaranteed.

Conflict of interest: Nil **Source of support:** None

References

- Çakmak C, Uğurluoğlu Ö. The effects of patient-centered communication on patient engagement, health-related quality of life, service quality perception and patient satisfaction in patients with cancer: A cross-sectional study in Türkiye. *Cancer Control* 2024;31:10732748241236327.
- Robinson JH, Callister LC, Berry JA, Dearing KA. Patient-centered care and adherence: Definitions and applications to improve outcomes. *J Am Acad Nurse Pract* 2008;20:600-7.
- Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, et al. Association of health literacy with diabetes outcomes. *JAMA* 2002;288:475-82.
- Braithwaite J, Runciman WB, Merry AF. Towards safer, better healthcare: Harnessing the natural properties of complex sociotechnical systems. *Qual Saf Health Care* 2009;18:37-41.
- Sekhar MS, Unnikrishnan Mk, Vyas N, Rodrigues GS. Development and evaluation of patient information leaflet for diabetic foot ulcer patients. *Int J Endocrinol Metab* 2017;15:e55454.
- Temsah A, Alhasan K, Altamimi I, Jamal A, Al-Eyadhy A, Malki KH, et al. DeepSeek in healthcare: Revealing opportunities and steering challenges of a new open-source artificial intelligence frontier. *Cureus* 2025;17:e79221.
- Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *J Med Internet Res* 2023;25:e46924.
- Jaques A, Abdelghafour K, Perkins O, Nuttall H, Haidar O, Johal K. A study of orthopedic patient leaflets and readability of AI-generated text in foot and ankle surgery (SOLE-AI). *Cureus* 2024;16:e75826.
- Haidar O, Jaques A, McCaughan PW, Metcalfe MJ. AI-generated information for vascular patients: Assessing the standard of procedure-specific information provided by the ChatGPT AI-language model. *Cureus* 2023;15:e49764.
- Daungsupawong H, Wiwanitkit V. Comment on: Assessing the quality and readability of online patient information: ENT UK patient information e-leaflets vs responses by a generative artificial intelligence. *Facial Plast Surg* 2025 Aug;41(4):472-481.
- Blom AW, Donovan RL, Beswick AD, Whitehouse MR, Kunutsor SK. Common elective orthopaedic procedures and their clinical effectiveness: Umbrella review of level 1 evidence. *BMJ* 2021 Jul 7;374:n1511.
- Edwards PK, Mears SC, Lowry Barnes C. Preoperative education for hip and knee replacement: Never stop learning. *Curr Rev Musculoskelet Med* 2017;10:356-64.
- Nash E, Bickerstaff M, Chetwynd AJ, Hawcutt DB, Oni L. The readability of parent information leaflets in paediatric studies. *Pediatr Res* 2023;94:1166-71.
- Abu-Heija AA, Shatta M, Ajam M, Abu-Heija U, Imran N, Levine D. Quantitative readability assessment of the internal medicine online patient information on annals.org. *Cureus* 2019;11:e4184.
- Hutchinson N, Baird GL, Garg M. Examining the reading level of internet medical information for common internal medicine diagnoses. *Am J Med* 2016;129:637-9.
- Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, et al. Readability of patient education materials from high-impact medical journals: A 20-year analysis. *J Patient Exp* 2021;8:2374373521998847.
- Crabtree L, Lee E. Assessment of the readability and quality of online patient education materials for the medical treatment of open-angle glaucoma. *BMJ Open Ophthalmol* 2022;7:e000966.
- Zrubka Z, Gulácsi L, Baji P, Kovács L, Tóth B, Fodor S, et al. eHealth Literacy and Health-Related Quality of Life: Associations between eHEALS, EQ-5D-5L and Health Behaviours in a Cross-Sectional Population Survey. Germany: Springer Science and Business Media LLC; 2023.
- Bostock S, Steptoe A. Association between low functional health literacy and mortality in older adults: Longitudinal cohort study. *Br Med J* 2012;344:e1602.
- AlSamarraie A, Househ M. The use of large language models in generating patient education materials: A scoping review. *Acta Inform Med* 2025;33:4-10.

Conflict of Interest: Nil
Source of Support: Nil

Consent: The authors confirm that informed consent was obtained from the patient for publication of this case report

How to Cite this Article

Jido JT, Wizni AA. Are Artificial Intelligence-generated Patient Leaflets Ready for Clinical Use? A Readability Comparison across Common Orthopaedic Procedures. *Journal of Orthopaedic Case Reports* 2025 October;15(10):295-301.

