

Artificial Intelligence (AI) -Assisted Optimization of Online Gastrointestinal Patient Education Materials: A Cross-Sectional Study

Vriti Khurana, Medha Tekriwal, Hany Eskarous

Submitted to: Journal of Medical Internet Research
on: October 08, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	17
Multimedia Appendixes	18
Multimedia Appendix 1.....	18

Preprint
JMIR Publications

Artificial Intelligence (AI) –Assisted Optimization of Online Gastrointestinal Patient Education Materials: A Cross-Sectional Study

Vriti Khurana¹; Medha Tekriwal² MS; Hany Eskarous³ MD

¹ St George's University School of Medicine, West Indies, Grenada Great River US

² Mellon College of Science, Carnegie Mellon University, Pittsburgh US

³ The Wright Center for Graduate Medical Education Scranton US

Corresponding Author:

Vriti Khurana

St George's University School of Medicine, West Indies, Grenada
3500 Sunrise Highway, Building 300
Great River
US

Abstract

Background: Patient education materials (PEMs) related to gastroenterology are often written at a level above the recommended sixth-grade reading level, which is suboptimal for accessibility. Large language models (LLMs) can enhance the readability of PEMs, but their effectiveness in this regard remains to be evaluated.

Objective: The goal of this research is to identify if LLMs can optimize the readability and understandability of gastroenterology (GI)-focused PEMs to a more accessible level, and if different models perform differently.

Methods: A cross-sectional review was performed on 60 PEMs that were randomly sampled from three GI-focused websites (American Cancer Society [ACS], American College of Gastroenterology [ACG], and American Gastroenterological Association [AGA]). PEMs were rewritten by four LLMs (ChatGPT, Gemini, Claude, and Perplexity) with a standardized fifth-grade translation prompt. Readability was assessed with the Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, and Simple Measure of Gobbledygook (SMOG) Index. Understandability was assessed with the Patient Education Materials Assessment Tool-Understandability (PEMAT-U). Accuracy was checked by two physicians who independently reviewed the simplified materials, with discrepancies verified using ChatGPT.

Results: PEMs from the original websites scored higher than the National Institutes of Health (NIH)-recommended sixth-grade level, with those from the ACG even at postgraduate levels. While all LLMs improved PEM readability and understandability, their performance and accuracy varied. Gemini had the most significant impact on readability but also produced the highest inaccuracy rate (11.6%). Claude introduced inaccuracies at a rate of 5%, while ChatGPT and Perplexity produced no errors. The accuracy review showed errors were concentrated in the most complex source materials and included oversimplification and omission of risk qualifiers.

Conclusions: LLMs show the potential to increase the accessibility of PEMs related to gastroenterology but vary in their performance, indicating the importance of human review. Gemini was the most effective of those included, but the inconsistency in performance and accuracy across different models suggests that AI output cannot be blindly trusted. A combined approach using LLMs and expert review can help improve patient understanding and health literacy.

(JMIR Preprints 08/10/2025:85486)

DOI: <https://doi.org/10.2196/preprints.85486>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/preprint/85486>, the full text will be available to the public.

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/preprint/85486>, the full text will be available to the public.



Original Manuscript

Artificial Intelligence (AI) –Assisted Optimization of Online Gastrointestinal Patient Education Materials: A Cross-Sectional Study

Authors: Vriti Khurana BS¹, Medha Tekriwal², Dr. Hany Eskarous MD³

¹St George's University School of Medicine, West Indies, Grenada

²Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA, mtekriwa@andrew.cmu.edu

³The Wright Center for Community Health, Scranton, PA

Abstract

Background: Patient education materials (PEMs) related to gastroenterology are often written at a level above the recommended sixth-grade reading level, which is suboptimal for accessibility. Large language models (LLMs) can enhance the readability of PEMs, but their effectiveness in this regard remains to be evaluated.

Objective: The goal of this research is to identify if LLMs can optimize the readability and understandability of gastroenterology (GI)-focused PEMs to a more accessible level, and if different models perform differently.

Methods: A cross-sectional review was performed on 60 PEMs that were randomly sampled from three GI-focused websites (American Cancer Society [ACS], American College of Gastroenterology [ACG], and American Gastroenterological Association [AGA]). PEMs were rewritten by four LLMs (ChatGPT, Gemini, Claude, and Perplexity) with a standardized fifth-grade translation prompt. Readability was assessed with the Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, and Simple Measure of Gobbledygook (SMOG) Index. Understandability was assessed with the Patient Education Materials Assessment Tool-Understandability (PEMAT-U). Accuracy was checked by two physicians who independently reviewed the simplified materials, with discrepancies verified using ChatGPT.

Results: PEMs from the original websites scored higher than the National Institutes of Health (NIH)-recommended sixth-grade level, with those from the ACG even at postgraduate levels. While all LLMs improved PEM readability and understandability, their performance and accuracy varied. Gemini had the most significant impact on readability but also produced the highest inaccuracy rate (11.6%). Claude introduced inaccuracies at a rate of 5%, while ChatGPT and Perplexity produced no errors. The accuracy review showed errors were concentrated in the most complex source materials and included oversimplification and omission of risk qualifiers.

Conclusions: LLMs show the potential to increase the accessibility of PEMs related to gastroenterology but vary in their performance, indicating the importance of human review. Gemini was the most effective of those included, but the inconsistency in performance and accuracy across different models suggests that AI output cannot be blindly trusted. A combined approach using LLMs and expert review can help improve patient understanding and health literacy.

Objective:

We sought to evaluate whether four large language models (LLMs)—ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic PBC), and Perplexity—can optimize the readability of patient education materials (PEMs) focused on gastrointestinal (GI) topics to a more accessible reading level.

Introduction:

Online health information is a primary resource for US adults seeking health advice, often used more than information from healthcare professionals, family, or traditional media [1,2]. While these online materials can be used alongside conversations between patients and physicians [3], many individuals use this information as their sole source for healthcare [1,2], which is concerning as this information is not tailored to individual needs and requires users to distinguish valuable information from irrelevant content [3,4,5]. With approximately 36% of US adults having only basic or below-basic health literacy skills [6], navigating the vast amount of online health information is a significant challenge [4,5].

Health literacy, defined as the ability to understand and use health information for decision-making [7], is crucial for patient outcomes. Individuals with low health literacy often experience poorer health outcomes [8] and may be more likely to have emergency department readmissions or use inappropriate health services [9]. Consequently, there is a growing recognition that creators of health information must present materials in a simplified format [10,11]. The National Institutes of Health (NIH) recommends that PEMs be written at a maximum of a sixth-grade reading level [12].

However, PEMs available online and in hospitals consistently fail to meet this standard, creating barriers for individuals with limited health literacy [13]. Studies focusing on specific diseases like stroke and heart failure have found materials written well above the recommended level [14]. This is a critical issue in gastroenterology, where patients must comprehend complex information about chronic conditions, procedures, and dietary changes [15]. This study focuses specifically on PEMs from prominent GI organizations, a context where clear communication is essential for effective patient self-management [16].

Large language models (LLMs), which can process and generate human-like text, offer an enticing solution to improve the readability of PEMs on a mass scale [17]. Previous studies have shown that LLMs can enhance the readability of discharge summaries and institution-specific PEMs but often require clinician oversight to ensure accuracy [18,19]. While patients or PEM creators could use publicly accessible LLMs like ChatGPT, Gemini, and Claude, little is known about the efficacy or safety of this approach for simplifying a broad range of health topics from major national organizations [20]. This study aims to fill that gap by evaluating whether these LLMs can effectively simplify online PEMs on GI topics from leading organizations, while also considering understandability.

Methods

Design & Sampling: Cross-sectional study analyzed 60 randomly selected online PEMs from three prominent websites that are commonly used for GI information: the American Cancer Society (ACS), the American College of Gastroenterology (ACG), and the American Gastroenterological Association (AGA). The selection of 60 articles was determined a priori based on feasibility and representation.

Intervention: We used four freely available generative artificial intelligence (AI) platforms: ChatGPT, Gemini, Claude, and Perplexity. Each of the 60 original PEMs was processed by each LLM using the prompt: "Translate the following text to a fifth-grade reading level." This prompt was chosen based on research showing that targeting a level lower than the National Institutes of Health (NIH) recommended sixth-grade level can account for variability in LLM outputs and improve results. This process resulted in five versions of each article: the original and four LLM-simplified

versions.

Outcomes: Primary outcome was the readability of the original online PEMs compared with the LLM-simplified versions. Readability was measured using four validated indices: the Flesch Reading Ease (Flesch) score, where a higher score on a 0-100 scale indicates easier readability; Flesch-Kincaid Grade Level (FKGL), which estimates the number of years of education a person needs to understand it, with a lower score meaning the text is simpler; the Gunning Fog Index (FOG), which also estimates the years of formal education needed to understand a text; and the Simple Measure of Gobbledygook (SMOG) index, which estimates the grade level required for comprehension, particularly for health literacy contexts [21,22,23]. The secondary outcome was to assess understandability using the Patient Education Materials Assessment Tool-Understandability (PEMAT-U), which scores materials based on clarity, organization, and absence of jargon [24].

Descriptive statistics, including means and standard deviations (SD), were calculated for all scores and word counts. A one-way ANOVA test was used to assess significant differences between the original PEMs from the three websites, and the Wilcoxon signed rank test was used to compare the LLM-simplified PEMs to their corresponding originals. The results are presented in Table 1.

A secondary outcome was the accuracy of the simplified materials. Assessments for inaccuracies were performed by two physicians and a medical student who independently reviewed the LLM-simplified text, comparing it to the original content as the criterion standard. Any identified discrepancies were then verified using ChatGPT to determine the final outcome. Errors were defined as instances in which the original message was changed, the content misrepresented the meaning of the original, or information was omitted, which could lead to a different interpretation by the reader. This definition is not strictly limited to factual medical inaccuracies but includes any change that alters the original meaning. The results are presented in Table 2.

Results

We noticed that one of the PEMs at ACG was an outlier with very high readability scores. Therefore, in Table 1, we have presented results for only 59 PEMs, excluding the outlier. However, in Table 3 in the appendix, we present the results including the outlier. The results in Table 1 show that the original PEMs from all three GI-focused websites were written at a reading level significantly above the recommended sixth-grade level. The ACG had the most complex materials, with a mean FOG score of 16.5 and a SMOG score of 14.6, indicating an almost postgraduate reading level.

Reading Score Analysis

- **American Cancer Society (ACS):** The original material had a FOG score of 12.6. Gemini was the most effective, lowering the FOG score to 9.3 ($p < 0.001$). Claude performed poorly, significantly increasing the FOG score to 15.6 ($p = 0.03$). All LLMs significantly improved understandability ($P \leq 0.001$).
- **American Gastroenterological Association (AGA):** The original material had a FOG score of 11.4. Gemini again showed the best performance, significantly reducing the FOG to 8.0 ($p < 0.001$). Claude again made the text substantially more complex, raising the FOG score to 18.5 ($p = 0.009$). All LLMs significantly improved PEMAT-U scores ($p < 0.001$).
- **American College of Gastroenterology (ACG):** The original material was extremely difficult, with a FOG score of 16.5. Gemini was the most

successful, dramatically lowering the FOG score to 7.8 ($p < 0.001$), bringing the text from a postgraduate level down to approximately an 8th-grade level. All LLMs significantly improved readability and understandability ($p \leq 0.005$).

LLM Performance: Gemini was overall the most effective, reducing ACG PEMs from postgraduate to middle school readability. ChatGPT and Perplexity yielded moderate but significant improvements. Claude increased grade level complexity in ACS and AGA materials, though it improved PEMAT-U scores.

Understandability: PEMAT-U scores improved significantly across all LLMs ($p < 0.005$).

Table 1. Mean reading scores and SD of the original version of PEM and each large language model version (N = 59).

Original		Claude		P value ^a	ChatGPT		P value ^a	Gemini		P value ^a	Perplex
Value, (SD)	mean	Value, (SD)	mean		Value, (SD)	mean		Value, (SD)	mean		Value, (SD)
57.9 (9.3)		53.4 (20.0)		0.21	62.8 (9.4)		<0.001	73.7 (9.3)		<0.001	68.2 (8.3)
9.8 (1.9)		13.0 (6.5)		0.02	8.6 (1.9)		<0.001	6.6 (1.8)		<0.001	7.4 (1.4)
12.6 (2.5)		15.6 (6.8)		0.03	11.2 (2.4)		<0.001	9.3 (2.1)		<0.001	9.9 (1.9)
11.9 (1.8)		12.6 (3.5)		0.25	10.8 (1.8)		<0.001	9.4 (1.7)		<0.001	9.9 (1.5)
75.6 (11.9)		84.5 (6.9)		<0.001	84.0 (7.0)		<0.001	85.0 (8.6)		<0.001	84.8 (6.3)
1296.3 (543.6)		836.6 (362.6)		<0.001	527.0 (112.4)		<0.001	342.7 (103.2)		<0.001	515.8 (103.2)
57.1 (9.0)		44.1 (24.8)		0.05	60.0 (12.9)		0.29	74.5 (5.4)		<0.001	64.7 (6.3)
9.1 (2.0)		16.2 (9.9)		0.008	9.4 (3.9)		0.73	6.1 (1.0)		<0.001	8.4 (1.8)
11.4 (2.2)		18.5 (10.2)		0.009	11.4 (4.1)		0.98	8.0 (1.0)		<0.001	10.4 (1.8)
11.2 (1.6)		13.3 (2.6)		0.005	10.8 (2.4)		0.49	8.5 (0.8)		<0.001	10.2 (1.6)
76.7 (9.1)		84.9 (5.9)		<0.001	87.7 (5.5)		<0.001	85.8 (7.3)		<0.001	86.0 (5.5)

Original	Claude		ChatGPT		Gemini		Perplex
957.6 (549.5)	703.4 (180.1)	0.02	412.1 (65.2)	<0.001	288.6 (41.1)	<0.001	412.2 (5
36.9 (7.8)	47.7 (15.3)	0.002	53.6 (9.7)	<0.001	75.7 (6.2)	<0.001	59.6 (10
13.7 (2.5)	14.5 (4.6)	0.41	10.2 (2.8)	<0.001	6.0 (1.0)	<0.001	9.3 (2.9
16.5 (2.5)	16.9 (4.7)	0.7	12.4 (2.9)	<0.001	7.8 (1.2)	<0.001	11.5 (2.
14.6 (1.5)	13.3 (2.6)	0.03	11.7 (1.8)	<0.001	8.2 (1.0)	<0.001	10.9 (1.
84.1 (9.6)	89.8 (5.6)	0.005	88.9 (7.5)	0.001	88.0 (7.1)	0.001	89.7 (5.
539.6 (383.2)	580.2 (169.5)	0.49	288.9 (136.3)	<0.001	232.6 (56.4)	0.001	334.8 (

^a *P* values reported are the large language model text compared with original source text, using Wilcoxon signed rank test.

^b Flesh: Flesch Reading Ease.

^c FKGL: Flesch-Kincaid Grade Level.

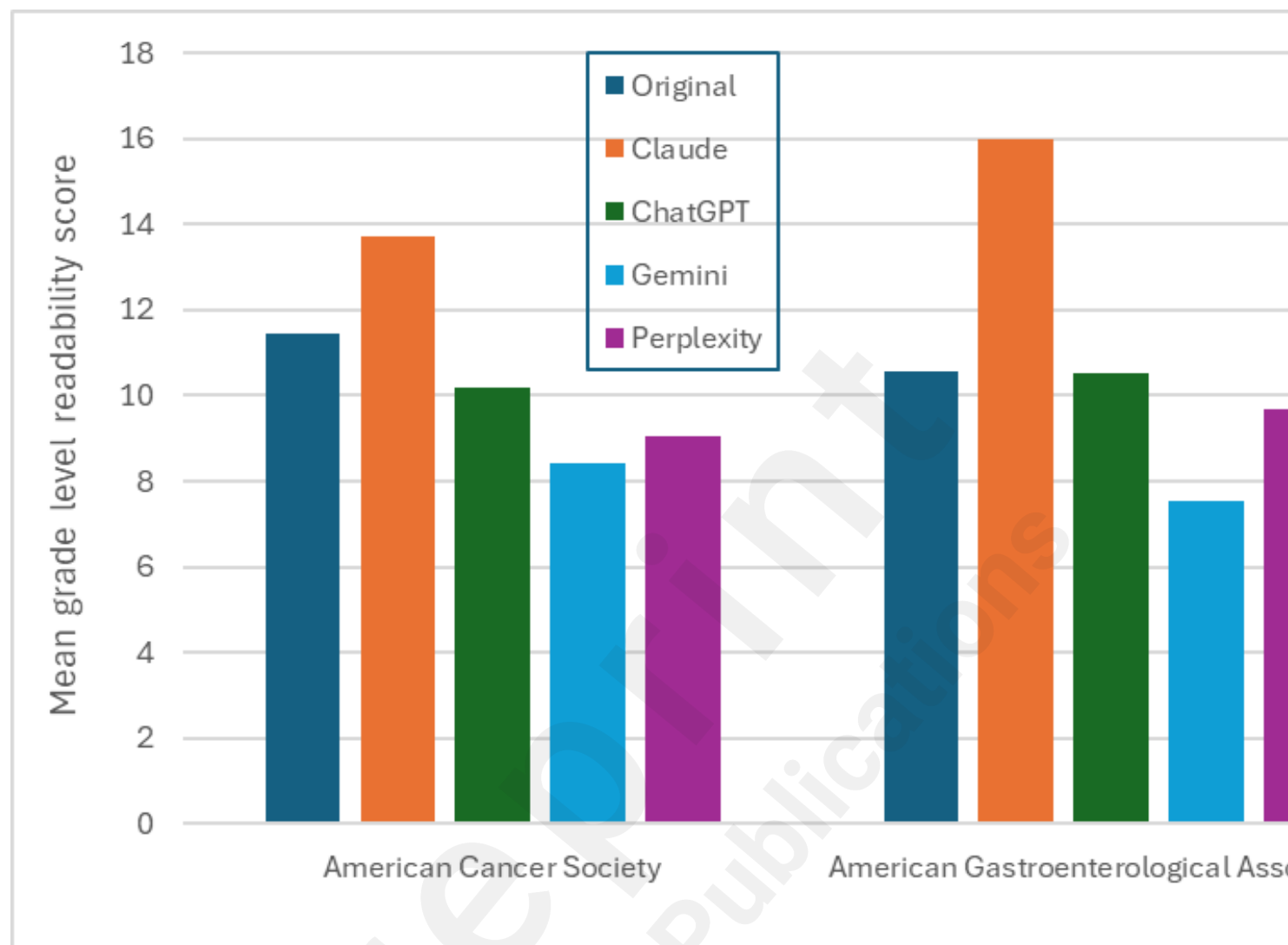
^d FOG: Gunning-Fog Index.

^e SMOG: Simple Measure of Gobbledygook Index,

^f PEMAT-U: Patient Education Materials Assessment Tools–Understandability.

We also calculated the average of the 3 grade-level readability scores (FKGL, FOG, and SMOG) for each PEM version of the material, and compared them across sites. Figure 1 shows mean readability scores by website and PEM version. This figure also shows that Gemini was the most effective for each of the websites.

Figure 1. Mean grade level scores for the original version of PEM and each large language model version (N=59).



Inaccuracy Data Analysis

Based on all 60 PEMs (including the outlier for ACG), Table 2 reveals significant differences in the accuracy of the four LLMs.

- ChatGPT and Perplexity were the most accurate, producing zero inaccuracies across all 60 articles.
- Gemini had the highest number of inaccuracies, with a total of 7 errors, for an overall inaccuracy rate of 11.6%. These errors were found in materials from the American Gastroenterological Association (2 errors, 10% rate) and the American College of Gastroenterology (5 errors, 25% rate).
- Claude produced 3 inaccuracies, for an overall rate of 5%. All of its errors occurred in materials from the American College of Gastroenterology (15% rate for that source).
- Notably, no inaccuracies were found in any of the simplified versions of materials from the American Cancer Society.

Table 2. Number of inaccuracies and inaccuracy rate of each large language model version when compared with the original.

Website	Claude, n (%)	ChatGPT n (%)	Gemini, n (%)	Perplexity, n (%)
American Cancer Society (N=20)	0 (0)	0 (0)	0 (0)	0 (0)
American Gastroenterological Association (N=20)	0 (0)	0 (0)	2 (10)	0 (0)
American College of Gastroenterology (N=20)	3 (15)	0 (0)	5 (25)	0 (0)
Total (N=60)	3 (5)	0 (0)	7 (11.6)	0 (0)

This analysis highlights a critical trade-off: Gemini, which proved to be the most effective model for improving readability, was also the least accurate. Conversely, the models that were perfectly accurate (ChatGPT and Perplexity) were less effective at simplification.

Discussion

This study's findings reaffirm a persistent issue in health communication: online patient education materials from leading gastroenterology organizations are written at levels far exceeding the accessibility standards for most U.S. adults [13,15]. Among the Large Language Models (LLMs) evaluated as a potential solution, Gemini proved to be the most powerful tool for enhancing readability [17]. It was uniquely capable of transforming the most complex materials, such as those from the American College of Gastroenterology (ACG) written at a postgraduate level, down to a more manageable eighth-grade reading level.

A key finding of this study is the apparent trade-off between simplification and accuracy. Gemini produced the largest gains in reading ease, but these improvements came at the expense of fidelity, with inaccuracies appearing in 11.6% of its outputs. In contrast, ChatGPT and Perplexity offered smaller improvements in readability while maintaining complete accuracy, with no errors detected in any of the simplified materials [17,18]. This pattern suggests that models capable of greater simplification may take more liberties with the original text, increasing the likelihood of introducing errors.

Claude's performance presented a notable paradox. For materials from the American Cancer Society (ACS) and the American Gastroenterological Association (AGA), the model often produced text that was more complex according to readability measures such as the Gunning Fog Index. At the same time, these outputs received higher scores on understandability assessments, including the PEMAT-U [16]. This discrepancy highlights the limitations of relying solely on automated readability metrics and suggests that users might mistakenly assume that any LLM-generated text is automatically easier to understand.

These results further underscore that current LLMs are not yet fully reliable for improving readability and should be applied with caution. In particular, errors in more technically complex materials demonstrate that expert human review remains essential to ensure both accuracy and patient comprehension [11].

While patients must approach such tools with caution, the greater responsibility lies with the authors and institutions that produce patient education resources. A balanced strategy—employing LLMs as preliminary drafting aids under the supervision of clinical experts—appears to offer the most effective route toward developing health information that is both accurate and accessible.

Limitations

We recognize several limitations in our approach. The analysis focused on content from three gastrointestinal health websites, and therefore, the results may not generalize across other sources. The analysis was further restricted to English-language texts; resources written in other languages, which often present additional challenges for readability and accuracy, were not included. Another limitation is the reliance on a single prompt. It is likely that different prompting strategies could have influenced the quality of the outputs. In addition, PEMAT-U scoring was performed manually, which introduces some degree of subjectivity despite efforts to apply consistent criteria. Finally, although we recorded how often inaccuracies occurred, we did not attempt to grade their clinical importance or potential effect on patient decision-making. Future studies could address these gaps by examining the severity of errors and developing structured methods to link readability gains with meaningful patient outcomes.

Conclusions

Patient education materials (PEMs) related to gastrointestinal health often exceed the sixth-grade reading level recommended by the National Institutes of Health (NIH) [12, 13,15]. This creates clear barriers for individuals with lower health literacy who may struggle to engage with these resources [6,8]. Large language models (LLMs) have been proposed as one possible solution, and certain models, such as Gemini, showed a marked ability to improve readability [11,17].

These findings, however, point to an important trade-off between readability and accuracy. Gemini produced the greatest gains in readability but also generated the most errors, with inaccuracies appearing in 11.6% of its outputs. These typically arose from oversimplification or omission of risk qualifiers and were concentrated in the more advanced source materials. Notably, the inaccuracy rate reached 25% when Gemini was applied to postgraduate-level texts from the American College of Gastroenterology (ACG). By contrast, ChatGPT and Perplexity were less effective in improving readability but did not introduce errors across the outputs we evaluated.

These results highlight the need for caution when utilizing AI to simplify text. Improvements in readability do not always guarantee that the information is accurate, and errors can still appear. For this reason, it remains essential for experts to review and verify content to ensure it is both accurate and comprehensible to patients. Future work should move beyond reporting error rates to examine the clinical significance of inaccuracies and to design structured approaches for integrating LLMs into the development of patient education materials. With careful application and oversight, these tools may eventually help produce PEMs that are both more accessible and clinically sound.

References

1. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, Viswanath K. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Arch Intern Med*. 2005;165(22):2618-2624. doi:10.1001/archinte.165.22.2618

2. Fox S, Duggan M. Health online 2013. *Pew Research Center*. Published January 15, 2013. Accessed October 2, 2025. <https://www.pewresearch.org/internet/2013/01/15/health-online-2013/>
3. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. *J Med Internet Res*. 2015;17(1):e3588. doi:10.2196/jmir.3588
4. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: Insights on the digital divide from the Health Information National Trends Survey 2012. *J Med Internet Res*. 2014;16(7):e172. doi:10.2196/jmir.3117
5. Beaunoyer E, Andrade N, Gonzalez M, Guitton MJ. Information and communication technologies (ICTs) and older adults: A critical review of the evidence. *Front Psychol*. 2017;8:1637. doi:10.3389/fpsyg.2017.01637
6. Kutner M, Greenberg E, Jin Y, Paulsen C. *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006-483)*. U.S. Department of Education, National Center for Education Statistics; 2006. <https://nces.ed.gov/pubs2006/2006483.pdf>
7. Institute of Medicine (US) Committee on Health Literacy. *Health Literacy: A Prescription to End Confusion*. National Academies Press; 2004. doi:10.17226/10883
8. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med*. 2011;155(2):97-107. doi:10.7326/0003-4819-155-2-201107190-00005
9. Griffey RT, Kennedy SK, McGowan L, Goodman M, Kaphingst KA. Is low health literacy associated with increased emergency department utilization and recidivism? *Acad Emerg Med*. 2014;21(10):1109-1115. doi:10.1111/acem.12488
10. Stableford S, Mettger W. Plain language: a strategic response to the health literacy challenge. *J Public Health Policy*. 2007;28(1):71-93. doi:10.1057/palgrave.jphp.3200102
11. Will J, et al. Enhancing the readability of online patient education materials: a large language model evaluation study. *J Med Internet Res*. 2025;27(1):e69955. doi:10.2196/69955
12. National Institutes of Health. *Clear & Simple: Developing Effective Print Materials for Low-Literate Readers*. U.S. Department of Health and Human Services; 2013. <https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/plain-language/clear-simple>
13. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. *Clin Orthop Relat Res*. 2010;468(10):2572-2580. doi:10.1007/s11999-010-1380-y
14. Ferguson LA. Low health literacy: implications for patient care. *J Am Acad Nurse Pract*. 2012;20(3):152-157. doi:10.1111/j.1745-7599.2012.00700.x

15. Kothari S, Moolla A, Bank S. Readability assessment of internet-based patient education materials related to gastroenterology. *Gastroenterol Rep (Oxf)*. 2018;6(2):132-138. doi:10.1093/gastro/gox032
16. Beaunoyer E, Dupéré S, Guitton MJ. Harnessing the digital divide in health information: a systematic review on the effects of the digital divide on health literacy and health disparities. *J Med Internet Res*. 2017;19(12):e429. doi:10.2196/jmir.6835
17. Thirunavukarasu AJ, Hassan A, Alonge O, Mahmud A. Can ChatGPT improve the readability of patient-facing materials? An evaluation of AI-assisted health education. *JMIR Med Educ*. 2023;9:e48296. doi:10.2196/48296
18. Jeblick K, Schachtner B, Dextl J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *Eur Radiol*. 2023;33:2943-2950. doi:10.1007/s00330-022-09252-4
19. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):e107-e108. doi:10.1016/S2589-7500(23)00023-7
20. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. doi:10.3390/healthcare11060887
21. Gunning R. *The Technique of Clear Writing*. McGraw-Hill; 1952.
22. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32(3):221-233. doi:10.1037/h0057532
23. McLaughlin GH. SMOG grading: a new readability formula. *J Reading*. 1969;12(8):639-646.
24. Shoemaker SJ, Wolf MS, Brach C. *The Patient Education Materials Assessment Tool (PEMAT) and User's Guide*. Agency for Healthcare Research and Quality; 2014. <https://www.ahrq.gov/health-literacy/patient-education/pemat.html>



Supplementary Files

Multimedia Appendixes

Untitled.

URL: <http://asset.jmir.pub/assets/094c835227450ea8e51ddcf75a426e61.doc>