MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

Assessing the readability of responses produced by ChatGPT and Gemini when answering questions about the gastrointestinal system

Okkeş Zortuk[1], Cihan Bedel [2]

1. Burdur Mehmet Akif Ersoy University, Faculty of Medicine, Burdur, Turkey
2. Antalya Research and Training Hospital, Emergency Medicine Department, Burdur, Turkey

ORCİD. https://orcid.org/0000-0001-6776-2702
ORCID. https://orcid.org/0000-0002-4473-9427

DOI: *https://doi.org/10.64288/4g35rv02*

Abstract

The utilization of artificial intelligence has proven to be a pivotal element in the timely identification of gastrointestinal diseases, thereby markedly enhancing the detection of lesions and ensuring enhanced diagnostic accuracy. A comparison of the AI models, ChatGPT-4.0 and Gemini, revealed distinct strengths and applications across various fields. Although AI can significantly advance gastrointestinal pharmacological research, broader implications and challenges must be considered. The objective of this study was to compare the responses of AI models to questions on gastrointestinal system pharmacology and readability.
This study was conducted using 30 multiple-choice questions in the field of pharmacology. The questions were answered and evaluated using two LLMs: ChatGPT-4.0, developed by Open AI, and Gemini 2.0, developed by Google. The analysis of readability and comprehensibility values in English was compared using the Automated Readability Index (ARI), Flesch-Kincaid, Gunning Fog Index, Coleman-Liau Index, SMOG score, and FORCAST scores. The average score for responses provided by ChatGPT-4.0 was 26.78±0.41, whereas the average score for responses provided by GEMINI was 28.90±0.91. The number of correct answers provided by GEMINI was significantly higher than that provided by O ChatGPT-4.0 (p=0.045). A readability comparison of the 30 questions was performed. The average ChatGPT-4.0 score for ARI was 13.04±1.77, whereas the average score for GEMINI was 14.76±2.04, and a significant difference was observed between them (p<0.001).

The present study demonstrated discrepancies in the utilization of gastrointestinal system pharmacology by ChatGPT-4.0 and Google Gemini, in addition to alterations in the readability of the responses.

Keywords: ChatGPT, Google Gemini, readability, artificial intelligence

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4g35rv02
E-ISSN: 3106-4418

MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

1. Introduction

Gastrointestinal system pharmacology encompasses a broad spectrum of pharmaceutical agents and therapeutic modalities directed towards the management and treatment of pathologies affecting the gastrointestinal (GI) tract. These include, but are not limited to, peptic ulcers, gastroesophageal reflux disease (GERD), irritable bowel syndrome (IBS), inflammatory bowel disease (IBD), and functional dyspepsia. Pharmacological approaches to these conditions involve various drug classes that target different aspects of GI physiology, such as secretion, motility, and inflammation [1,2]. Recent advancements in the field of artificial intelligence (AI) have precipitated the notable integration of this technology into the domain of gastroenterology. This integration has resulted in substantial progress in the diagnosis, management, and treatment of GI diseases. The utilization of artificial intelligence (AI) technologies, particularly machine learning and deep learning, has demonstrated potential in enhancing diagnostic accuracy, optimizing patient management, and facilitating the early detection of GI disorders. This integration is transforming traditional practices in gastroenterology and providing more efficient and accurate approaches to patient care [3,4]. The utilization of artificial intelligence has proven to be a pivotal element in the timely identification of gastrointestinal diseases, thereby markedly enhancing the detection of lesions and ensuring enhanced diagnostic accuracy. This is particularly evident in the use of AI for the analysis of endoscopic and radiological images, where AI systems have been trained to differentiate between benign and malignant lesions with a high degree of accuracy [5].

A comparison of the AI models, ChatGPT 4.0 and Gemini 2.0, reveals distinct strengths and applications across various fields, including healthcare, business management, and research. Both models are capable of providing accurate information and enhancing operational efficiency. However, they differ in their specific strengths and application areas [6]. The ensuing discourse aims to meticulously unravel the comparative performance, applications, and limitations of these technologies by drawing upon relevant research papers [7]. In the medical field, Gemini has been shown to outperform ChatGPT-4.0 in terms of the accuracy and comprehensiveness of its responses. For instance, in the context of sudden sensorineural hearing loss (SSHL), Gemini demonstrated higher levels of completion and accuracy, although the difference in accuracy was not statistically significant. In the field of urology, Gemini demonstrated proficiency in the identification of congenital penile curvature, whereas ChatGPT-4.0 exhibited a particular aptitude in the formulation of management strategies for renal artery aneurysms [8]. However, both models demonstrated deficiencies in diagnostic accuracy and the occurrence of hallucinations. Gemini demonstrated a higher level of accuracy in its responses when compared to ChatGPT, particularly in the context of multiple-choice questions [9]. However, ChatGPT-4.0 exhibited superior

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

performance on open-ended and true/false questions. In a recent study, ChatGPT 4.0 demonstrated superior performance in the diagnosis of complex hematologic cases, both in terms of primary and differential diagnoses, when compared to Gemini Advanced [10].

Although AI can significantly advance gastrointestinal pharmacological research, its broader implications and challenges must be considered. Integrating AI into healthcare requires careful consideration of data privacy, model interpretability, and robust validation across diverse populations.

This study aimed to evaluate the effectiveness of advanced AI models, such as ChatGPT 4.0, in improving diagnostic accuracy in complex hematologic cases and to explore the broader implications, challenges, and considerations involved in integrating AI technologies into gastrointestinal system pharmacology research and healthcare practice, with particular attention to data privacy, model interpretability, and validation across diverse populations.

2. Methods

This study was conducted by three experts in the field of pharmacology using 30 questions with answers. The questions were prepared using "Katzung's Basic & Clinical Pharmacology" as a reference [11]. The questions were designed by experts in groups of ten as easy, medium, and difficult. The questions were answered and evaluated using ChatGPT-4.0, developed by ChatGPT-4.0, and Gemini 2.0, developed by Google, both of which are large language models. As a result of the observation that this particular study did not involve the use of human subjects, animals, or living materials, the requirement for ethical committee approval was disregarded in this instance.

Accuracy Comparison

To compare the correct answers given to the questions by the LLMs, 30 questions were reanswered on 20 different days with the browser's cookies reset. The order in which the questions were asked varied.

Readability Comparison

An analysis of the ease and difficulty of reading and understanding English was conducted. The Average Reading Level Consensus calculates the average reading level by averaging other scales.

The automated readability index (ARI) is a measure that assesses the readability of a text. Although opinions differ regarding its accuracy compared to syllable/word and complex word indices, the character/word index is generally calculated faster because character counts are easier and more accurate for computer programs than syllable counts. In fact, this index was designed for the real-time monitoring of the readability of electric typewriters [12].

$$4.71\left(\frac{\text{characters}}{\text{words}}\right)+0.5\left(\frac{\text{words}}{\text{sentences}}\right)-21.43$$

Flesch-Kincaid readability tests are readability tests designed to show how difficult an English text is to understand. The "Flesch-Kincaid" (F-K) reading level was

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

developed in 1975 by J. Peter Kincaid and his team under a contract with the US Navy. In the Flesch readability test, higher scores indicate material that is easier to read, whereas lower scores indicate text that is more difficult to read. The formula for the Flesch readability score (FRES) test is as follows [13,14]:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

In linguistics, the **Gunning Fog Index** is a readability test for English writing. The index estimates the number of years of formal education required for a person to understand a text on the first reading. For example, a Fog Index of 12 requires a reading level of a high school senior in the United States (approximately 18 years old). The test was developed in 1952 by Robert Gunning, an American businessman active in newspaper and textbook publications. Furthermore, the Fog Index is primarily valid for English and may not accurately reflect readability in other languages [15].

- Select a passage of approximately 100 words (e.g., one or more complete paragraphs). Do not skip any sentences.
- Determine the average sentence length. (Divide the number of words by the number of sentences.);
- Count "complex" words consisting of three or more syllables. Do not include proper nouns, familiar jargon, or compound words. Do not count common suffixes (e.g., -es, -ed, or -ing) as syllables.
- Add the average sentence length and the percentage of complex words; and

The Coleman-Liau index is a readability test designed by Meri Coleman and T. L. Liau to measure the comprehensibility of a text. Similar to ARI, but unlike most other indices, the Coleman-Liau index is based on characters rather than syllables per word. The Coleman-Liau index is designed to be easily calculated mechanically from printed text samples. The Coleman-Liau index is calculated using the following formula [16]

$$:CLI = 0.0588 \cdot L - 0.296 \cdot S - 15.8$$

The SMOG score is a readability measure that estimates the years of education required to understand a text. SMOG stands for "Simple Measure of Gobbledygook." The SMOG index has no statistical validity for languages other than English. SMOG formula [17,18]:

$$\text{grade} = 1.0430\sqrt{\text{number of polysyllables} \times \frac{30}{\text{number of sentences}}} + 3.1291$$

Linsear Write is a readability measure for English texts, allegedly developed to help the U.S. The standard Linear Write metric, Lw, operates on a 100-word sample: The standard Linsear Write metric, Lw, operates on a 100-word sample:

- One point was awarded for each "eawotorootorto defineword with two or fewer syllables.
- Three points were awarded for each "difficult word" defined as a word with three or more syllables.

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

• Divide the points by the number of sentences in the 100-word sample.

• Sprovidesnal resuIf if r > 20, then Lwr/2/ 2. If r ≤ 20, then Lw = r / 2 - 1.

The result is a "grade level" measure that reflects the estimated years of schooling required to read the text fluently [19].

The FORCAST note is a readability measure that indicates how difficult a text is to read based on the number of single-syllable "easy" words in a sample of 100–150 words. The values correspond to the number of years of schooling required to understand the text. For example, six years corresponds to readers aged 11–12, while 12 years corresponds to readers aged 17–18 [20].

Statistical Analysis

To evaluate the findings obtained in the study, the Statistical Package for the Social Sciences (SPSS) for Windows 27.0 program was used for statistical analyses. The data were then classified. Categorical data are defined as percentages and frequencies. Numerical data are defined, and distribution analysis is performed. Data that conform to a normal distribution are defined as the mean ± standard deviation (SD). Parametric tests (t-test and analysis of variance [ANOVA]) were used in the analysis of numerical tests that conform to a normal distribution. Subsequent analysis of the data was conducted using an unpaired t-test. The mean of repeated measures was calculated for each dataset. Findings with a p-value of < 0.05 were considered significant.

3. Results

The average score for responses provided by ChatGPT-4.0 was 26.78±0.41, whereas that for responses provided by GEMINI was 28.90±0.91. The number of correct answers provided by GEMINI was significantly higher than that provided by ChatGPT-4.0 (p=0.045). The score comparisons are shown in Table 1. A readability comparison was performed for the 30 questions. The average ChatGPT-4.0 score for ARI was 13.04±1.77, whereas the average score for GEMINI was 14.76±2.04, and a significant difference was observed between them (p<0.001). The readability comparisons are shown in Table 2.

Table 1: Score comparisons

	Open AI (n=30)	GEMINI (n=30)	p-Value
Easy	8,78±0,78	9,50±0,51	0,002**
Modarate	9,00±0,66	9,65±0,48	0,001**
Hard	9,00±0,00	9,75±0,44	<0,001*
Total	26,78±0,41	28,90±0,91	0,045*

*1-tailed t-test, **2-tailed t-test

Table 2: Readability comparisons

	Open AI (n=30)	GEMINI (n=30)	p-Value
ARLCalc	13,41±1,29	13,98±1,05	0,23*
ARI	13,04±1,77	14,76±2,04	<0,001**
Flesch Reading Ease	24,07±12,13	24,53±7,86	0,121*
Gunning Fog Index	15,39±2,49	15,42±1,51	0,028*
Flesch-Kincaid Grade Level	13,27±1,93	14,57±1,71	0,008**
Coleman-Liau Readability Index	15,12±2,09	14,90±1,25	0,137*

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

The SMOG Index	11,31±0,97	12,97±1,45	0,003*
Original Linsear Write Formula	63,90±8,24	59,00±5,23	<0,001*
Linsear Write Grade Level Formula	12,23±2,81	13,33±2,95	0,338*
FORCAST Readability Formula	12,83±0,69	12,15±0,29	<0,001*

*1-tailed t-test, **2-tailed t-test

3. Discussion

The integration of artificial intelligence (AI) into the field of gastroenterology holds considerable promise for the diagnosis, treatment, and management of gastrointestinal (GI) diseases, with the potential for significant transformation. Recent advancements in AI technologies, with a particular emphasis on machine learning (ML) and convolutional neural networks (CNNs), have demonstrated considerable potential for enhancing diagnostic accuracy, improving patient outcomes, and optimizing clinical workflows. This integration is particularly impactful in areas such as endoscopy, pathology, and pharmacology, where AI is capable of processing and analyzing large datasets with greater precision than traditional methods [2-5]. The utilization of artificial intelligence (AI) has proven to be a pivotal element in the timely and accurate diagnosis of gastrointestinal diseases. This technological advancement has facilitated the precise identification of lesions and cancerous alterations, thereby contributing to the effective management and treatment of these conditions. For instance, AI systems have been developed to differentiate between benign and malignant lesions by analyzing endoscopic and radiological images, with the capacity to achieve optimal diagnostic outcomes [6,21]. In endoscopy, AI-driven image analysis has enhanced the detection of conditions such as Barrett's esophagus and esophageal squamous cell carcinoma, often outperforming human endoscopists in terms of accuracy and speed [22].

The integration of artificial intelligence (AI) models, such as ChatGPT and Gemini, in the field of gastrointestinal system pharmacology offers a promising avenue for enhancing both educational and clinical applications. These AI tools have been evaluated for their capacity to generate pharmacology-related content, assist in medical inquiries, and support pharmacometric tasks. However, the effectiveness of these devices varies across different applications, necessitating further refinement and expert oversight to maximize their utility in gastrointestinal pharmacology [20-22]. ChatGPT has been employed to generate multiple-choice questions for pharmacology education, demonstrating its capacity to adhere to structural guidelines and provide educational content. Nevertheless, ensuring medical accuracy and comprehensiveness remains challenging, as both are pivotal for reliable utilization in medical education [23]. In the context of gastrointestinal pharmacology, ChatGPT has been employed to generate questions and explanations for examinations, with findings demonstrating moderate to high agreement in terms of content accuracy and clinical relevance. However, issues with the cognitive level and quality of distractors have been identified, suggesting the need for

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

MedEpicent
Journal of Medical Education and Clinical Research

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

expert review and improvement [24]. The evaluation of ChatGPT and Gemini was conducted to ascertain their capacity to generate non-model evaluation method (NONMEM) codes for pharmacometric tasks pertinent to clinical pharmacology settings. While these templates can provide a valuable starting point, the output frequently exhibits errors that require correction by experienced professionals [25]. In the domain of gastrointestinal diseases, the efficacy of ChatGPT in accurately diagnosing prevalent conditions, including irritable bowel syndrome and inflammatory bowel disease, has been assessed. The model has demonstrated the capacity to improve patient education and physician–patient communication. Nevertheless, its function as a tool to educate physicians requires further investigation [26]. In the field of gastroenterology, ChatGPT has demonstrated superior performance in terms of accuracy and reliability compared to Google Bard, particularly in medical management tasks. This lends credence to its potential as a reliable instrument in the field of study, although further research and development are necessary to enhance its capabilities [27]. In the present study, a higher number of correct answers were provided by GEMINI than those provided by ChatGPT-4.0. This discrepancy should be considered when assessing the applicability of AI technologies in pharmacology.

The readability and effectiveness of AI chatbots, such as ChatGPT and Google Gemini, in generating content for various medical and educational purposes have been subjects of extensive research. The prevailing focus of these studies is on the readability, accuracy, and appropriateness of information provided by AI models [28]. The following section presents the results of the readability scores. Google Gemini has been shown to generally produce content that is more legible than that generated by ChatGPT. For instance, in the context of emergency medical conditions, Gemini's brochures were slightly more accessible than those of ChatGPT, with a higher ease score. Similarly, in the domain of refractive surgery FAQs, Gemini was noted for its relatively superior readability, although all chatbots require a university-level understanding [29]. The present study examined the relationship between content quality and sentence counts in emergency medical brochures. Nevertheless, Gemini demonstrates a particular aptitude in producing succinct and legible responses, a skill that is of paramount importance for ensuring patient comprehension [29,30]. ChatGPT frequently demonstrates superior accuracy, particularly in intricate medical scenarios, such as retinal detachment and intraoperative decision support in plastic surgery. Nevertheless, Gemini has been observed to provide more suitable responses in specific contexts, including frequently asked questions (FAQs) in refractive surgery. In terms of appropriateness, Gemini demonstrated superior performance in providing suitable responses to FAQs concerning refractive surgery, thereby signifying its capacity to deliver contextually relevant information [30]. The potential of both AI models in the field of medical education has been rigorously evaluated. ChatGPT demonstrated marginally superior accuracy and comprehensiveness in patient

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

education materials for local anesthesia in eye surgery. Conventional patient information leaflets (PILs) retain their overall superior performance [31]. In the context of end-of-life care, Google Gemini was demonstrated to have superior readability and actionability, although both models conveyed positive sentiments and high levels of accuracy [32]. In the present study, the mean ChatGPT-4.0 score for ARI was 13.04 ± 1.77, while the mean score for GEMINI was 14.76 ± 2.04. A significant difference was observed between them, which may be an important result in the context of evaluating the readability of AI models.

Our study had some limitations. First, it used only two AI tools. Better results can be obtained by evaluating other AI tools. Inclusion of a greater number of diseases would have provided greater clarity. Second, as chatbots are frequently updated, the use of an older version of AI may be necessary. In some cases, AI may not be able to provide up-to-date medical information because it may be difficult to access.

4. Conclusion

The present study demonstrated discrepancies in the utilization of gastrointestinal system pharmacology by ChatGPT and Google Gemini, in addition to alterations in the readability of the responses.

Acknowledgements

Access to the artificial intelligence applications utilized in this study was established on August 20, 2025. During this process, Chat GPT 4.0, developed by OpenAI, and Gemini 2.0, developed by Google, were utilized. The authors declare no conflicts of interest among themselves or with other institutions. No financial support was received for this study.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

5. References

1. Bjarnason I, Scarpignato C, Holmgren E, Olszewski M, Rainsford KD, Lanas A. Mechanisms of Damage to the Gastrointestinal Tract From Nonsteroidal Anti-Inflammatory Drugs. Gastroenterology. 2018 Feb;154(3):500-514. doi: 10.1053/j.gastro.2017.10.049. Epub 2017 Dec 6. PMID: 29221664..
2. Sharkey KA, Mawe GM. The enteric nervous system. Physiol Rev. 2023 Apr 1;103(2):1487-1564. doi: 10.1152/physrev.00018.2022. Epub 2022 Dec 15. PMID: 36521049; PMCID: PMC9970663.
3. Bunne C, Roohani Y, Rosen Y, Gupta A, Zhang X, Roed M, Alexandrov T, AlQuraishi M, Brennan P, Burkhardt DB, Califano A, Cool J, Dernburg AF, Ewing K, Fox EB, Haury M, Herr AE, Horvitz E, Hsu PD, Jain V, Johnson GR, Kalil T, Kelley DR, Kelley SO, Kreshuk A, Mitchison T, Otte S, Shendure J, Sofroniew NJ, Theis F, Theodoris CV, Upadhyayula S, Valer M, Wang B, Xing E, Yeung-Levy S, Zitnik M, Karaletsos T, Regev A, Lundberg E, Leskovec J, Quake SR. How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities. ArXiv [Preprint]. 2024 Oct 14:arXiv:2409.11654v2. Update in: Cell. 2024 Dec 12;187(25):7045-7063. doi:

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

10.1016/j.cell.2024.11.015. PMID: 39398201; PMCID: PMC11468656.

4. Kleinstreuer N, Hartung T. Artificial intelligence (AI)-it's the end of the tox as we know it (and I feel fine). Arch Toxicol. 2024 Mar;98(3):735-754. doi: 10.1007/s00204-023-03666-2. Epub 2024 Jan 20. PMID: 38244040; PMCID: PMC10861653.

5. Pecere S, Milluzzo SM, Esposito G, Dilaghi E, Telese A, Eusebi LH. Applications of Artificial Intelligence for the Diagnosis of Gastrointestinal Diseases. Diagnostics (Basel). 2021 Aug 30;11(9):1575. doi: 10.3390/diagnostics11091575. PMID: 34573917; PMCID: PMC8469485.

6. Is EE, Menekseoglu AK. Comparative performance of artificial intelligence models in rheumatology board-level questions: evaluating Google Gemini and ChatGPT-4o. Clin Rheumatol. 2024 Nov;43(11):3507-3513. doi: 10.1007/s10067-024-07154-5. Epub 2024 Sep 28. PMID: 39340572.

7. Ryan DK, Maclean RH, Balston A, Scourfield A, Shah AD, Ross J. Artificial intelligence and machine learning for clinical pharmacology. Br J Clin Pharmacol. 2024 Mar;90(3):629-639. doi: 10.1111/bcp.15930. Epub 2023 Nov 12. PMID: 37845024.

8. Abdala Kuri S, Morales C, Oliva AM, Peña A, Dévora S. Analysis of the level of polypharmacy in patients from an isolated rural area: effect of age, sex, and chronic diseases. Front Digit Health. 2025 Jun 10;7:1508505. doi: 10.3389/fdgth.2025.1508505. PMID: 40556639; PMCID: PMC12185413.

9. Mayer B, Kringel D, Lötsch J. Artificial intelligence and machine learning in clinical pharmacological research. Expert Rev Clin Pharmacol. 2024 Jan;17(1):79-91. doi: 10.1080/17512433.2023.2294005. Epub 2024 Jan 23. PMID: 38165148.

10. Singh H, Nim DK, Randhawa AS, Ahluwalia S. Integrating clinical pharmacology and artificial intelligence: potential benefits, challenges, and role of clinical pharmacologists. Expert Rev Clin Pharmacol. 2024 Apr;17(4):381-391. doi: 10.1080/17512433.2024.2317963. Epub 2024 Feb 15. PMID: 38340012.

11. Vanderah TW (2024). In: Katzung's Basic & Clinical Pharmacology, 16th Edition. McGraw-Hill, New York, NY,

12. Smith EA, Senter RJ (1967) Automated readability index. AMRL-TR Aerospace Medical Research Laboratories (US):1-14

13. Kincaid JP, Braby R, Mears JE (1988) Electronic authoring and delivery of technical information. Journal of instructional development 11:8-13. https://doi.org/10.1007/BF02904998

14. Kincaid JP, Aagard JA, Hara JWO, Cottrell LK (1981) Computer readability editing system. IEEE Transactions on Professional Communication PC-24:38-42. https://doi.org/10.1109/TPC.1981.6447821

15. Świeczkowski D, Kułacz S (2021) The use of the Gunning Fog Index to evaluate the readability of Polish and English drug leaflets in the context of Health Literacy challenges in Medical Linguistics: An exploratory study. Cardiology journal 28:627-631. https://doi.org/10.5603/CJ.a2020.0142

16. Coleman Liau Index Calculator. (2025) Calculator Doc. https://calculatordoc.com/coleman-liau-index-calculator. 2025

17. Fitzsimmons PR, Michael BD, Hulley JL, Scott GO (2010) A readability assessment of online Parkinson's disease information. The journal of the Royal College of Physicians of

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4g35rv02
E-ISSN: 3106-4418

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

Edinburgh 40:292-296. https://doi.org/10.4997/jrcpe.2010.401
18. Contreras A, García-Alonso R, Echenique M, Daye-Contreras F (1999) The SOL formulas for converting SMOG readability scores between health education materials written in Spanish, English, and French. Journal of health communication 4:21-29. https://doi.org/10.1080/108107399127066
19. Duffy TM (1985) CHAPTER 6 - Readability Formulas: What's the Use?*. In: Duffy TM, Waller R (eds) Designing Usable Texts. Academic Press, pp 113-143. https://doi.org/https://doi.org/10.1016/B978-0-12-223260-2.50011-6
20. Oliffe M, Thompson E, Johnston J, Freeman D, Bagga H, Wong PKK (2019) Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional Health Literacy Study. BMJ open 9:e024582. https://doi.org/10.1136/bmjopen-2018-024582
21. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. World J Gastroenterol. 2022 Jan 7;28(1):108-122. doi: 10.3748/wjg.v28.i1.108. PMID: 35125822; PMCID: PMC8793013.
22. Koleth G, Emmanue J, Spadaccini M, Mascagni P, Khalaf K, Mori Y, Antonelli G, Maselli R, Carrara S, Galtieri PA, Pellegatta G, Fugazza A, Anderloni A, Selvaggio C, Bretthauer M, Aghemo A, Spinelli A, Savevski V, Sharma P, Hassan C, Repici A. Artificial intelligence in gastroenterology: Where are we heading? Endosc Int Open. 2022 Nov 15;10(11):E1474-E1480. doi: 10.1055/a-1907-6569. PMID: 36397868; PMCID: PMC9666060.
23. Bedel HA, Bedel C, Selvi F, Zortuk Ö, Karancı Y. Emergency Medicine Assistants in the Field of Toxicology, Comparison of ChatGPT-3.5 and GEMINI Artificial Intelligence Systems. Acta Med Litu. 2024;31(2):294-301. doi: 10.15388/Amed.2024.31.2.18. Epub 2024 Dec 4. PMID: 40060265; PMCID: PMC11887820.
24. Laohawetwanit T, Apornvirat S, Kantasiripitak C. ChatGPT as a teaching tool: Preparing pathology residents for board examination with AI-generated digestive system pathology tests. Am J Clin Pathol. 2024 Nov 4;162(5):471-479. doi: 10.1093/ajcp/aqae062. PMID: 38795049.
25. Shin E, Yu Y, Bies RR, Ramanathan M. Evaluation of ChatGPT and Gemini large language models for pharmacometrics with NONMEM. J Pharmacokinet Pharmacodyn. 2024 Jun;51(3):187-197. doi: 10.1007/s10928-024-09921-y. Epub 2024 Apr 24. PMID: 38656706.
26. Kerbage A, Kassab J, El Dahdah J, Burke CA, Achkar JP, Rouphael C. Accuracy of ChatGPT in Common Gastrointestinal Diseases: Impact for Patients and Providers. Clin Gastroenterol Hepatol. 2024 Jun;22(6):1323-1325.e3. doi: 10.1016/j.cgh.2023.11.008. Epub 2023 Nov 19. PMID: 37984563.
27. Fattah FH, Salih AM, Salih AM, Asaad SK, Ghafour AK, Bapir R, Abdalla BA, Othman S, Ahmed SM, Hasan SJ, Mahmood YM, Kakamad FH. Comparative analysis of ChatGPT and Gemini (Bard) in medical inquiry: a scoping review. Front Digit Health. 2025 Feb 3;7:1482712. doi: 10.3389/fdgth.2025.1482712. PMID: 39963119; PMCID: PMC11830737.
28. Rouhi AD, Ghanem YK, Yolchieva L, Saleh Z, Joshi H, Moccia MC, Suarez-Pierre A, Han JJ. Can Artificial Intelligence Improve the Readability of Patient Education Materials on Aortic Stenosis? A Pilot Study. Cardiol Ther. 2024 Mar;13(1):137-147. doi: 10.1007/s40119-023-00347-0. Epub 2024 Jan 9. PMID: 38194058; PMCID: PMC10899139.
29. S A, Aggarwal S, Sridhar J, Vs K, John VP, Singh C. An Observational Study to Evaluate

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4g35rv02
E-ISSN: 3106-4418

Received: 9 February 2026
Revised: 18 March 2026
Accepted: 20 March 2026
Published: 31 March 2026

Readability and Reliability of AI-Generated Brochures for Emergency Medical Conditions. Cureus. 2024 Aug 31;16(8):e68307. doi: 10.7759/cureus.68307. PMID: 39350844; PMCID: PMC11441454.

30. Aydın FO, Aksoy BK, Ceylan A, Akbaş YB, Ermiş S, Kepez Yıldız B, Yıldırım Y. Readability and Appropriateness of Responses Generated by ChatGPT 3.5, ChatGPT 4.0, Gemini, and Microsoft Copilot for FAQs in Refractive Surgery. Turk J Ophthalmol. 2024 Dec 31;54(6):313-317. doi: 10.4274/tjo.galenos.2024.28234. PMID: 39743925; PMCID: PMC11707452.

31. Gondode P, Duggal S, Garg N, Lohakare P, Jakhar J, Bharti S, Dewangan S. Comparative Analysis of Accuracy, Readability, Sentiment, and Actionability: Artificial Intelligence Chatbots (ChatGPT and Google Gemini) versus Traditional Patient Information Leaflets for Local Anesthesia in Eye Surgery. Br Ir Orthopt J. 2024 Aug 19;20(1):183-192. doi: 10.22599/bioj.377. PMID: 39183761; PMCID: PMC11342839.

32. Gondode PG, Khanna P, Sharma P, Duggal S, Garg N. End-of-life Care Patient Information Leaflets-A Comparative Evaluation of Artificial Intelligence-generated Content for Readability, Sentiment, Accuracy, Completeness, and Suitability: ChatGPT vs Google Gemini. Indian J Crit Care Med. 2024 Jun;28(6):561-568. doi: 10.5005/jp-journals-10071-24725. PMID: 39130387; PMCID: PMC11310687.

Vol. 2 No. 1 (2026)
Orginal Article
DOI: 10.64288/4935rv02
E-ISSN: 3106-4418