



## Comparative Analysis of the Readability of Artificial Intelligence-Supported Educational Materials for Acute Cholecystitis Patients: ChatGPT and Gemini

By

Muhammed ali Erinmez

Department of Emergency Medicine, Mersin Provincial Health Directorate, Mersin, Türkiye



### Article History

Received: 15/05/2026

Accepted: 28/05/2026

Published: 30/05/2026

### Vol – 5 Issue –5

PP: - 75-77

### Abstract

**Introduction:** Acute cholecystitis is a common gastrointestinal disease characterized by inflammation of the gallbladder, and patient education constitutes an important component of the treatment process. In recent years, artificial intelligence (AI)-based language models have increasingly been used in the preparation of patient education materials. However, the readability level of these materials is critically important for patients to understand health information. This study aimed to compare the readability of acute cholecystitis educational materials generated by ChatGPT and Gemini.

**Methods:** In this comparative study, AI-generated patient education texts about acute cholecystitis were analyzed. Twenty texts were obtained from OpenAI ChatGPT and twenty from Gemini. Readability was evaluated using the Average Reading Level Consensus, Automated Readability Index (ARI), Flesch Reading Ease, Fog Scale, Flesch–Kincaid Grade Level, Coleman–Liau Index, SMOG Index, Linsear Write Formula, and Forcast Readability Formula. Differences between groups were analyzed statistically, and  $p < 0.05$  was considered significant.

**Results:** Texts generated by Gemini demonstrated higher readability in the Flesch Reading Ease score ( $58.48 \pm 3.42$  vs.  $44.28 \pm 4.36$ ;  $p < 0.001$ ). Fog Scale results also revealed that Gemini content was easier to read ( $11.44 \pm 1.42$  vs.  $15.48 \pm 1.36$ ;  $p < 0.001$ ). In terms of the Automated Readability Index, ChatGPT showed lower scores ( $9.32 \pm 2.38$  vs.  $12.42 \pm 2.44$ ;  $p = 0.032$ ). No significant differences were found in other readability indices.

**Conclusion:** Both AI models produced readable texts that could be used for patient education regarding acute cholecystitis. However, Gemini provided more understandable content in some important readability measures. Although AI-supported patient education materials have the potential to improve health literacy, readability should be optimized for broader patient populations.

**Keywords:** Acute cholecystitis, Artificial intelligence, Readability, ChatGPT, Gemini

### Introduction

Acute cholecystitis is one of the most common inflammatory diseases of the gastrointestinal system and a major cause of emergency hospital admissions worldwide. The disease is generally characterized by inflammation of the cystic duct of the gallbladder due to obstruction by gallstones [1,2]. Patients with acute cholecystitis frequently experience abdominal pain, fever, nausea, and gastrointestinal symptoms. Early diagnosis and appropriate treatment are crucial to prevent serious complications such as perforation, abscess, and sepsis. In

addition to medical and surgical treatment, patient education is a vital part of the treatment process [3,4]. Effective patient education improves adherence to treatment, reduces patient anxiety, and contributes to improved health outcomes [5,6].

In recent years, rapid advancements in artificial intelligence (AI) technologies, particularly the use of large language models in healthcare, have increased. AI-based systems like ChatGPT and Gemini enable the rapid and accessible preparation of patient education materials [7,8]. These systems allow for the quick production of informative content

for patients and facilitate access to health information. However, the effectiveness of these materials largely depends on their readability level. Complex and difficult-to-understand health texts can negatively impact health literacy and reduce active participation in the treatment process by making it difficult for patients to comprehend information [9,10].

Readability is a key measure that expresses how easily a written text can be understood by its target reader. Various validated readability scales such as the Flesch Reading Ease, Flesch–Kincaid Grade Level, SMOG Index, and Automated Readability Index are used for this purpose. The literature generally suggests that patient education materials should be prepared at a sixth-grade reading level to ensure they are understandable to a wide audience [11]. Therefore, evaluating the readability of AI-generated patient education content is crucial for the reliable use of these technologies in health communication. Although studies examining the medical applications of AI systems are increasing, studies evaluating the readability of AI-assisted patient education materials related to acute cholecystitis are limited [12]. The aim of this study is to compare the readability levels of acute cholecystitis patient education texts generated by ChatGPT and Gemini using different validated readability indices.

## Method

In this study, two commonly used artificial intelligence-based language models (OpenAI ChatGPT and Google Gemini) were used. Texts for patient education were generated based on the topic of acute cholecystitis (gallbladder inflammation). A total of 40 patient education texts, 20 generated by ChatGPT and 20 by Gemini, were included in the dataset for analysis. This comparative study evaluated the readability levels of patient education materials generated by two different artificial intelligence-based language models, ChatGPT and Gemini, regarding acute cholecystitis. A total of 40 educational texts were included in the study. 20 of these were generated by OpenAI ChatGPT and 20 by Gemini. All texts were prepared using standardized prompts to generate informative content for patients about acute cholecystitis. The generated content was analyzed as follows: The texts contained general information about the definition, symptoms, diagnostic methods, treatment options, complications, and prevention methods of acute cholecystitis.

The readability levels of the texts were evaluated using validated readability indices commonly used in health communication research. These indices included the Average Reading Level Consensus, Automated Readability Index (ARI), Flesch Reading Ease Score, Gunning Fog Scale, Flesch–Kincaid Grade Level, Coleman–Liau Index, SMOG Index, Linsear Write Formula, Dale–Chall Readability Score, and Spache Readability Formula. These readability formulas determine the comprehensibility level of the texts by analyzing different linguistic features such as sentence length, word complexity, syllable count, and word difficulty level.

All generated texts were evaluated electronically using readability analysis tools. The numerical data obtained were expressed as mean  $\pm$  standard deviation. The readability

scores of the texts generated by ChatGPT and Gemini were statistically compared. Before analysis, the normality of the data distribution was assessed, and appropriate statistical tests were used according to the distribution characteristics. The statistical significance level was accepted as  $p < 0.05$ .

## Results

A total of 40 patient education texts related to acute cholecystitis were analyzed, including 20 texts generated by ChatGPT and 20 texts generated by Gemini. The readability analyses demonstrated differences between the two artificial intelligence models across several readability indices. Gemini-generated texts showed significantly higher readability according to the Flesch Reading Ease score compared with ChatGPT-generated texts ( $58.48 \pm 3.42$  vs.  $44.28 \pm 4.36$ ;  $p < 0.001$ ). In addition, the Gunning Fog Scale results indicated that Gemini texts were easier to read and required a lower educational level for comprehension than ChatGPT texts ( $11.44 \pm 1.42$  vs.  $15.48 \pm 1.36$ ;  $p < 0.001$ ). In terms of the Automated Readability Index (ARI), ChatGPT-generated texts demonstrated lower scores compared with Gemini-generated texts ( $9.32 \pm 2.38$  vs.  $12.42 \pm 2.44$ ;  $p = 0.032$ ), indicating variability between readability formulas. No statistically significant differences were identified between the two models regarding the Coleman–Liau Index, SMOG Index, Linsear Write Formula, Dale–Chall Readability Score, or Spache Readability Formula ( $p > 0.05$  for all comparisons). Detailed comparisons of readability scores between ChatGPT and Gemini are presented in Table 1. Overall, both artificial intelligence models generated patient education materials at moderate readability levels; however, Gemini generally produced more comprehensible content according to several major readability measures.

## Discussion

This study compared the readability levels of patient education materials related to acute cholecystitis generated by two widely used artificial intelligence-based language models, ChatGPT and Gemini. The findings demonstrated that both AI systems were capable of producing readable and informative educational content; however, significant differences were observed in several readability indices [13,14]. In particular, Gemini-generated texts showed better readability performance according to the Flesch Reading Ease and Gunning Fog Scale scores, suggesting that these materials may be easier for patients to understand [15,16].

Health literacy is a critical determinant of patient outcomes, treatment adherence, and participation in medical decision-making. Educational materials that are difficult to read may reduce patients' understanding of their disease and negatively affect communication between healthcare professionals and patients [17,18]. Current recommendations suggest that patient education materials should ideally be written at approximately a sixth-grade reading level to ensure accessibility for broader populations [19,20]. Although both ChatGPT and Gemini produced texts above this recommended level, Gemini appeared to generate relatively simpler and more comprehensible content in some important

readability measures. The differences observed between readability formulas may be related to the distinct linguistic structures and text-generation strategies used by each AI model [21,22]. While some indices evaluate sentence length and word complexity more heavily, others focus on syllable count and vocabulary familiarity. For example, the lower Fog Scale scores observed in Gemini-generated texts may indicate the use of shorter sentences and less complex terminology. In contrast, ChatGPT-generated texts may have included more detailed medical explanations and advanced vocabulary, which could increase readability scores despite improving informational depth [23-25].

The growing use of artificial intelligence in healthcare communication offers important advantages, including rapid content production, accessibility, and scalability [26-28]. AI-generated educational materials may reduce the workload of healthcare professionals and provide patients with instant access to medical information. However, readability and accuracy remain essential considerations before integrating such technologies into routine patient education practices. Materials that are not sufficiently understandable may contribute to misinformation or confusion, especially among individuals with limited health literacy [29,30].

This study has several limitations. First, the analysis was limited to texts related to acute cholecystitis and may not be

generalizable to other medical conditions. Second, readability formulas evaluate linguistic characteristics but do not directly measure patient comprehension or educational effectiveness. Third, the study only compared two AI language models, and future studies including additional platforms and larger datasets may provide broader insights.

### Conclusion

Both ChatGPT and Gemini were capable of generating readable patient education materials related to acute cholecystitis. However, Gemini demonstrated superior performance in several important readability indices, particularly in terms of reading ease and lower educational level requirements. These findings suggest that AI-based language models may serve as useful tools for preparing patient-oriented educational content, although differences between models should be considered.

The readability of patient education materials is essential for improving health literacy, strengthening patient participation in the treatment process, and enhancing communication between healthcare professionals and patients. Therefore, optimization of AI-generated medical content for readability remains an important requirement before widespread clinical implementation.

	GEMINI (n=40)	CHAT –GPT (n=40)	p-Value
ARI	8,25±1,21	9.44±1,03	0.223
Flesch Reading Ease	55,44±2,45	56,62±4,12	0.112
Fog Scale (Gunning FOG Formula)	14.62±2,1	12.44±2.4	0.248
Flesch-Kincaid Grade Level	8.55±2.62	9,25±1.64	<0.001
Coleman-Liau Readability Index	11.84±3.12	15.62±1,82	0.244
The SMOG Index	8.56±2.22	9.44±2,01	0.148
Linsear Write Formula	60.12±10.660	66.38±10.84	0.332
Dale-Chall Readability Score	10.44±0.66	11.84±0.98	0.642
Spache Readability Formula	8.56±0.88	8.86±0.74	0.442

**Table 1. Comparison of CHATGPT and GEMINI in terms of readability scores.**