

AI Chatbots Deliver Readability Comparable to OrthoInfo for ACL Injury Education

Auston R Locke, Matthew Dylan Ramey, Srivatsan J Swaminathan, Niklas Koehne, Jonathan J Huang, Salvatore Capotosto, Thomas Boucher, James N Gladstone

INTRODUCTION: Patients increasingly rely on the internet for health-related information (HRI), particularly for common musculoskeletal conditions such as anterior cruciate ligament (ACL) injuries. These injuries often impact the careers of young athletes and the lives of patients, requiring surgical intervention, the consideration of numerous appealing techniques, and prolonged rehabilitation. In this complex landscape, the internet offers an overwhelming volume of patient-directed resources—many of which are inconsistent or controversial—leading to potential confusion. While professionally authored platforms like the American Academy of Orthopaedic Surgeons (AAOS) OrthoInfo offer validated educational content, the recent proliferation of artificial intelligence (AI) tools has raised questions about their reliability and readability. This study evaluates the readability of AI-generated content about ACL injuries and compares it with corresponding information from OrthoInfo.

METHODS: Six questions about ACL injuries and treatment were derived based on information from OrthoInfo’s page on ACL injuries. These questions covered topics including ACL anatomy, description, injury cause, injury symptoms, physical examinations, and injury treatment. ChatGPT-4 and Google Gemini 2.0 Flash were used to generate AI responses to these questions. Responses were generated on May 3rd, 2025 in incognito browser sessions. Corresponding information from OrthoInfo was used as a control. The readability of information was quantified using the Flesch-Kincaid Reading Ease Index, Coleman-Liau Index, Flesch-Kincaid Grade Level Index, FORCAST Readability Formula, Gunning Fog Index, and Simple Measure of Gobbledygook Index. Statistical testing was conducted using the Kruskal-Wallis nonparametric One-Way Analysis of Variance (ANOVA) test with significance indicated at $p < 0.05$.

RESULTS:

No statistically significant differences in mean readability scores were observed across specific content areas, including ACL anatomy, description, injury cause, symptoms, physical examination, and treatment (p -values: 0.075, 0.135, 0.725, 0.284, 0.736, and 0.832, respectively) (Table 1). When all content areas were pooled, mean readability scores favored OrthoInfo ($p = 0.004$), with OrthoInfo averaging 12.2 and both ChatGPT and Gemini averaging 12.4 (Table 1). The highest and lowest individual readability scores were both produced by Gemini: ACL anatomy content scored 13.5, while symptom-related content scored 10.9 (Table 1). Among readability formulas, the Flesch-Kincaid Grade Level Index yielded the lowest scores, while the Coleman-Liau Index produced the highest (Table 2, Figure 1). None of the health-related information (HRI) sources included in this study met the sixth-grade reading level recommended by the American Medical Association (AMA) and the National Institutes of Health (NIH).

DISCUSSION AND CONCLUSION: AI-generated content from ChatGPT and Gemini demonstrates readability levels comparable to those of OrthoInfo for ACL-related patient education. While statistical significance favored OrthoInfo when all questions were pooled, this difference is not clinically meaningful: OrthoInfo scored, on average, only 0.2 grade levels lower than the chatbots—a difference that would not be noticeable to patient readers. Notably, and alarmingly, all readability scores were well above the thresholds of the 6th grade level recommended by the American Medical Association (AMA) and National Institutes of Health (NIH). This highlights the need to simplify educational contents. In summary, clinicians should continue to guide patients toward trusted sources while helping to contextualize and simplify information. At the same time, they should recognize the evolving role of AI in delivering accessible health information.

Figure 1. Bar chart of grade level readability scores for OrthoInfo, ChatGPT, and Gemini, grouped by metric. Each bar shows the mean score from each scoring system for the six questions used in the study with corresponding lines representing standard deviation.

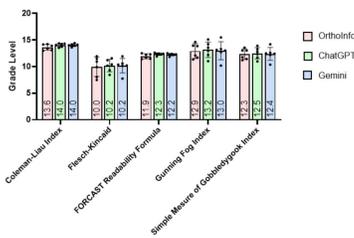


Table 1. Raw readability scores for OrthoInfo, ChatGPT, and Gemini across five OrthoInfo-sourced questions. All scores represent grade-level needed for comprehension, except for the Flesch-Kincaid reading ease which operates on an inverse scale.

Readability Metric	Anatomy		Description		Cause		Symptoms		Physical Examinations		Treatment		p-value						
	O	C	O	C	O	C	O	C	O	C	O	C							
Flesch-Kincaid Reading Ease Index	68.4	46.6	44.8	66.4	45.4	50.7	51.0	49.0	47.0	43.7	57.1	59.5	47.4	50.4	47.9	47.2	47.5	47.5	0.863
Coleman-Liau Index (Grade Level)	13.0	13.5	13.6	13.1	13.7	13.9	14.2	14.1	14.2	14.2	14.3	14.4	13.9	14.4	14.1	13.4	14.2	14.0	0.284
Flesch-Kincaid (Grade Level)	7.5	11.2	11.9	8.0	11.6	10.1	9.9	10.4	10.5	11.9	8.4	7.7	10.9	9.3	10.2	11.5	10.2	10.4	0.998
FORCAST Readability Formula (Grade Level)	11.4	12.6	12.2	11.5	12.3	12.2	12.5	12.4	12.3	11.7	11.9	11.8	12.3	12.4	12.5	12.2	12.4	12.4	0.243
Gunning Fog Index (Grade Level)	11.3	15.1	15.5	11.7	14.2	13.0	12.1	13.6	13.3	13.9	11.0	10.2	13.9	12.2	12.7	14.6	13.0	13.1	0.967
Simple Measure of Gobbledygook Index	11.2	13.0	14.2	11.4	13.3	12.4	11.8	12.8	12.6	13.0	11.0	10.4	13.0	11.7	12.2	13.8	12.3	12.5	0.984

Table 2. Mean and standard deviation of readability scores for OrthoInfo, ChatGPT, and Gemini across six OrthoInfo-sourced questions.

Category	Mean±SD	p-value		
		O	C	G
Anatomy	10.9±2.0	13.3±1.5	13.5±1.5	0.075
Description	11.1±1.9	13.0±1.1	12.4±1.4	0.137
Cause	12.1±1.5	12.6±1.4	12.6±1.3	0.725
Symptoms	13.0±1.1	11.3±2.1	10.9±2.4	0.284
Physical Examination	12.8±1.2	12.0±1.8	12.4±1.4	0.736
Treatment	13.1±1.2	12.4±1.5	12.5±1.3	0.832
All Questions	12.2±1.7	12.4±1.6	12.4±1.7	0.004*

Abbreviations: C, ChatGPT; G, Google Gemini; O, OrthoInfo; SD, Standard Deviation
* denotes significance via One-Way ANOVA among OrthoInfo, ChatGPT, and Gemini.