Supplementary Table 1: Comparison of Readability Grade Levels in Texts Generated by Artificial Intelligence Models							
Grade Level Category	**ChatGPT (n)**	**DeepSeek (n)**	**Gemini (n)**	**p ***	**ChatGPT vs DeepSeek p †**	**ChatGPT vs Gemini p †**	**DeepSeek vs Gemini p †**
ARI Grade				**0.002**	0.031	**0.005**	0.334
< 12th Grade	17	5	2				
12th Grade	14	16	12				
College	59	6	76				
ARI Difficulty				**0.002**	0.031	**0.005**	0.334
Slightly Difficult	4	0	0				
Somewhat Difficult	5	0	0				
Fairly Difficult	8	5	2				
Difficult	14	16	12				
Very Difficult	59	69	76				
ARI Age				**0.002**	0.031	**0.005**	0.334
14-18 yrs.	31	21	14				
18-22 yrs.	59	69	76				
FRE Level				**0.009**	0.046	0.078	**0.011**
Very Confusing	6	13	15				
Difficult	79	66	74				
Fairly Difficult	4	11	1				
Standard	1	0	0				
GFI Grade				**0.003**	0.817	**<0.001**	**0.007**
< 12th Grade	83	80	75				
12th Grade	4	5	12				
College	3	5	3				
FKGL Grade				**<0.001**	**<0.001**	**0.005**	**<0.001**

< 12th Grade	37	68	20				
12th Grade	24	12	24				
College	29	10	46				
SMOG Grade				**<0.001**	**<0.001**	**<0.001**	**<0.001**
< 12th Grade	16	56	0				
12th Grade	16	19	11				
College	57	15	79				
FORCAST Grade				**<0.001**	**<0.001**	**0.002**	**0.002**
< 12th Grade	41	8	17				
12th Grade	45	60	67				
College	4	22	6				
***** Chi-square							
† Bonferroni-corrected Z-test (the significance level was set at $p < 0.0167$)							
Abbreviations: ARI = Automated Readability Index; FRES = Flesch Reading Ease Score; GFI = Gunning Fog Index; FKGL = Flesch-Kincaid Grade Level; SMOG = Simple Measure of Gobbledygook; FORCAST = FORCAST Readability Formula							

Supplementary Table 2: Comparison of Texts Generated by Artificial Intelligence Models According to Reliability and Quality Metrics							
Metric	**ChatGPT (n)**	**DeepSeek (n)**	**Gemini (n)**	**p ***	**ChatGPT vs DeepSeek p †**	**ChatGPT vs Gemini p †**	**DeepSeek vs Gemini p †**
M-DISCERN				**< 0.001**	**< 0.001**	**< 0.001**	0.808
1	34	0	0				
2	27	28	27				
3	29	28	32				
4	0	34	31				
JAMA Benchmark				**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
0	50	0	26				
1	40	47	36				
2	0	43	28				
EQIP				**< 0.001**	0.622	**< 0.001**	**< 0.001**
Serious problems with quality	34	28	5				
Good quality with minor problems	53	58	76				
Well written	3	3	9				
GQS				**< 0.001**	**0.001**	**< 0.001**	0.873
2	30	10	8				
3	34	40	40				
4	26	40	42				

* Chi-square

† Pairwise comparisons for significant differences in categorical distributions were assessed using the Bonferroni correction (significance level set at $p < 0.0167$).

Abbreviations: M-DISCERN = Modified DISCERN; JAMA = Journal of the American Medical Association; EQIP = Ensuring Quality Information for Patients; GQS = Global Quality Score.