Further results

Table 6: LM scores per prompt for RQ1.

Model	Which are the criteria for making a text dyslexia friendly?	Which are the principles in Dyslexia Style guide that suggests changes in written material for dyslexic readers?	What guidelines should be followed to make a text more accessible for people with dyslexia?	How can I make a text dyslexia-friendly?
Gemma	10	10	11	9,5
Phi4	14	14	14,50	13
Gpt-4-turbo	13,5	13	13	12,5

Table 7: LMs average performance improvements from the original texts. In **bold**, we highlight the best improvement scores per model for the *DysText* visual, content, and total performance.

	Criteria Given											
	No guide prompts from RQ2						Full Guide			Text criteria		
	Visual	Content	Total	Visual	Content	Total	Visual	Content	Total	Visual	Content	Total
Gemma	**1.54**	**3.21**	**4.75**	0.26	2.59	2.85	0.80	1.75	2.55	1.38	2.45	3.83
Phi4	1.52	**2.64**	4.16	0.52	2.40	2.92	1.78	2.41	4.19	**1.96**	2.24	**4.20**
GPT4-turbo	0.64	1.60	2.24	0.18	1.68	1.86	1.84	1.33	3.17	**2.16**	**1.94**	**4.10**

Table 8: LMs percentage of improved *DysText* scores. Corresponding to Table 7.

	Criteria Given											
	No guide prompts from RQ2						Full Guide			Text criteria		
	Visual	Content	Total	Visual	Content	Total	Visual	Content	Total	Visual	Content	Total
Gemma	86%	100%	100%	16%	100%	100%	52%	94%	94%	96%	100%	100%
Phi4	86%	100%	100%	28%	100%	100%	100%	100%	100%	100%	100%	100%
GPT4-turbo	34%	98%	98%	16%	98%	98%	94%	92%	98%	100%	98%	98%

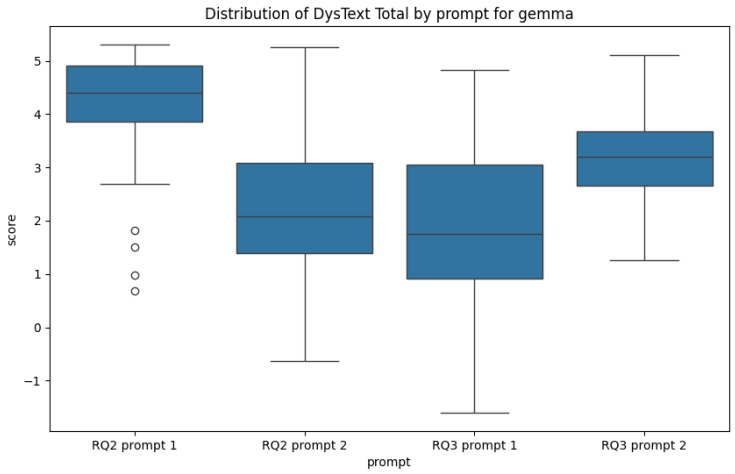

Fig. 2: Gemma box plot of prompts.

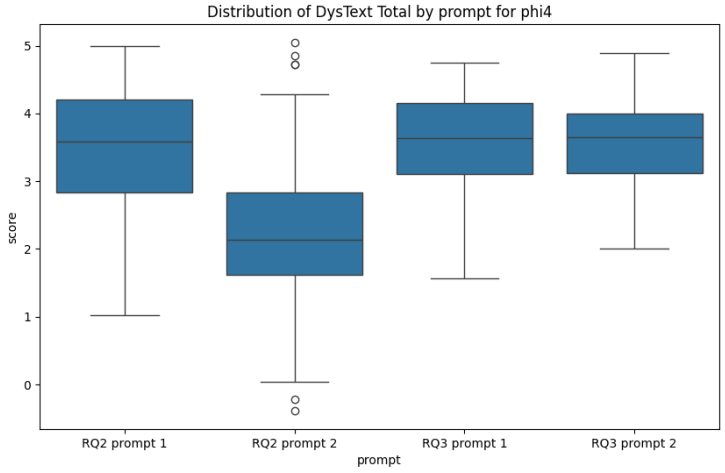

Fig. 3: Phi4 box plot of prompts.

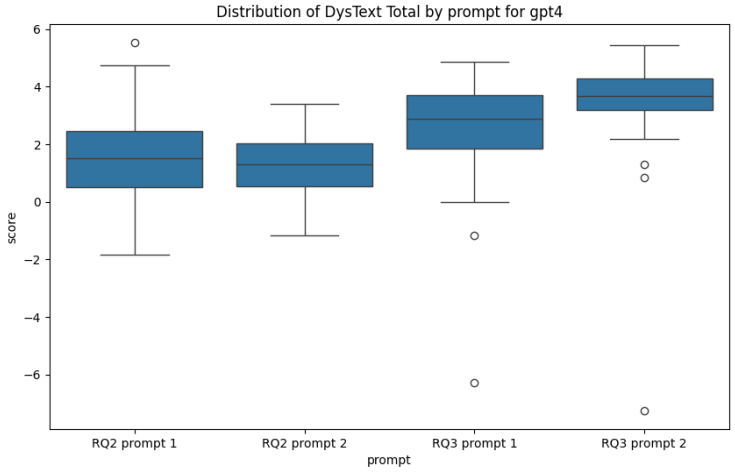

Fig. 4: GPT4-turbo box plot of prompts.