$p \le 0.001$ (	***)	
-----------------	------	--

		Sign Test Results Between Prompts																			
	random1	_	-0.50	-1.00	-1.00	-0.75	+0.67	-0.50	+0.67	+0.00	-0.75	-0.67	+0.67	-0.83	-0.80	-1.00*	-0.50	-0.60	-0.88	-0.86	-0.85*
	random2 -	0.50	_	-1.00	-1.00	-0.75	+0.67	-0.50	+0.67	-0.50	-0.75	-0.67	+0.67	-0.75	-0.80	-0.88	-0.50	-0.60	-0.80	-0.78	-0.80*
	random3	+1.00	+1.00	_	+1.00	+0.80	+1.00*	+0.78	+1.00*	+1.00	+0.67	+0.83	+0.88	+0.57	+0.62	-0.57	+0.86	+0.75	-0.56	-0.50	-0.62
	random4	+1.00	+1.00	-1.00	_	-0.50	+0.80	+0.67	+0.80	+1.00	-0.50	+0.60	+0.80	-0.67	-0.60	-0.83	+0.75	+0.57	-0.75	-0.71	-0.77
	random5 ·	+0.75	+0.75	-0.80	-0.50	_	+0.80	+0.75	+0.80	+0.75	-0.50	+0.60	+1.00	-0.67	-0.67	-0.83	+1.00	+0.60	-0.75	-0.71	-0.77
	semantic_random1	0.67	-0.67	-1.00*	-0.80	-0.80	_	-0.67	+0.00	-0.67	-0.80	-1.00	-0.50	-0.86	-0.83	-0.89*	-0.67	-0.75	-1.00*	-1.00*	-0.86*
	semantic_random2 -	0.50	-0.50	-0.78	-0.67	-0.75	+0.67	_	+0.67	-0.50	-1.00	-0.60	+1.00	-0.75	-0.80	-0.80	-0.50	-0.60	-1.00*	-1.00	-0.91*
	semantic_random3 -	0.67	-0.67	-1.00*	-0.80	-0.80	+0.00	-0.67	_	-0.67	-0.80	-1.00	-0.50	-0.86	-0.83	-0.89*	-0.67	-0.75	-1.00*	-1.00*	-0.86*
	semantic_random4 -	+0.00	-0.50	-1.00	-1.00	-0.75	+0.67	-0.50	+0.67	_	-0.75	-0.67	+0.67	-0.83	-0.80	-1.00*	-0.50	-0.60	-0.88	-0.86	-0.85*
ηρt A	semantic_random5	+0.75	+0.75	-0.67	-0.50	-0.50	+0.80	+1.00	+0.80	+0.75	_	+0.60	+1.00	-0.60	-0.60	-0.70	+1.00	+0.57	-0.75	-0.71	-0.77
Program	nantic_top_to_bottom ·	+0.67	+0.67	-0.83	-0.60	-0.60	+1.00	+0.60	+1.00	+0.67	-0.60	_	+0.75	-0.67	-0.75	-0.78	+0.67	-0.50	-0.78	-0.75	-0.79
sen	nantic_bottom_to_top -	0.67	-0.67	-0.88	-0.80	-1.00	-0.50	-1.00	-0.50	-0.67	-1.00	-0.75	_	-0.86	-1.00	-0.89*	-1.00	-0.75	-1.00*	-1.00*	-0.92**
	thematic_random1	+0.83	+0.75	-0.57	+0.67	+0.67	+0.86	+0.75	+0.86	+0.83	+0.60	+0.67	+0.86	_	+0.57	-0.67	+0.75	+0.80	-0.62	-0.60	-0.69
	thematic_random2 -	+0.80	+0.80	-0.62	+0.60	+0.67	+0.83	+0.80	+0.83	+0.80	+0.60	+0.75	+1.00	-0.57	_	-0.71	+1.00	+0.67	-0.67	-0.67	-0.80
	thematic_random3 -	+1.00*	+0.88	+0.57	+0.83	+0.83	+0.89*	+0.80	+0.89*	+1.00*	+0.70	+0.78	+0.89*	+0.67	+0.71	_	+0.88	+1.00	-0.50	+0.56	-0.64
	thematic_random4 -	0.50	-0.50	-0.86	-0.75	-1.00	+0.67	-0.50	+0.67	-0.50	-1.00	-0.67	+1.00	-0.75	-1.00	-0.88	_	-0.60	-0.88	-0.86	-0.85*
	thematic_random5 -	+0.60	+0.60	-0.75	-0.57	-0.60	+0.75	+0.60	+0.75	+0.60	-0.57	-0.50	+0.75	-0.80	-0.67	-1.00	+0.60	_	-0.78	-0.75	-0.79
the	matic_top_to_bottom -	+0.88	+0.80	+0.56	+0.75	+0.75	+1.00*	+1.00*	+1.00*	+0.88	+0.75	+0.78	+1.00*	+0.62	+0.67	-0.50	+0.88	+0.78	_	+0.67	-0.62
the	matic_bottom_to_top -	+0.86	+0.78	-0.50	+0.71	+0.71	+1.00*	+1.00	+1.00*	+0.86	+0.71	+0.75	+1.00*	+0.60	+0.67	-0.56	+0.86	+0.75	-0.67	_	-0.70
	deepseek_solutions	+0.85*	+0.80*	+0.62																	
	, and	domit	John?	John's no	Jornal Jornal	join's	joinit end	John 2 John	don's	don'd	John's by	Storn X	atic rand	jorn <sup>1</sup>	John?	John's not spend	Jona dictor	Jon's pr	Storn Y	to?	tions
					emai	ema	ena	eno	ena	i' x'	dicy	rhem	inem'	"hem	" hem	"hem	o X	dic	) Leeps	,~	

Positive value: Prompt A outperforms Prompt B Negative value: Prompt B outperforms Prompt A

Value represents proportion of disagreements  $* p \le 0.05$ ,  $** p \le 0.01$ ,  $*** p \le 0.001$  † Significant after Bonferroni correction

Prompt B

 $p \le 0.01 (**)$ 

 $p \le 0.05 (*)$ 

 $^{L}p > 0.05$