	Κ.
--	----

l-p	≤	0.01	(**)

 $p \le 0.05 (*)$ 

-0.50 + 0.60

-0.50

-0.60

-0.56

semantic top to bottom semantic bottom to top thenatic top to bottom thematic bottom to top deepseex solutions semantic randoma thenatic randoma thematic randoms thematic random? thematic random3 semantic Yandom's thematic randomiz

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

random1 -

random2

random3

random4

random5

-0.75 -0.80 -0.83 -0.50 -0.50 +1.00

semantic\_random3 - -0.69 -0.75 -0.73 -0.57 -0.60 +0.57 -0.60

-0.56

randomis

randoma

randomiz

 Semantic\_random5 - 0.78
 -0.83
 -0.86
 -0.60
 +0.67
 -1.00
 -0.50
 -0.50

 Semantic\_top\_to\_bottom - 0.78
 -0.78
 -0.83
 -0.86
 -0.60
 -0.60
 +0.60
 -0.67
 -0.50
 -0.50
 -0.50

thematic\_random4 - -0.50 +0.57 -0.50 +0.83 +0.83 +0.88 +0.83 +0.78 +0.78 +0.86 +0.78 +0.70 +0.75

semantic random?

semantic Jandomil

semantic random3

-0.75 -0.57 -0.75 +0.67 +0.67 +0.70 +0.62 +0.64 +0.64 +0.67 +0.67 +0.58

-0.67 -0.57 -0.67 +0.62 +0.75 +0.83 +0.75 +0.67 +0.71 +0.80 +0.71 +0.62 -0.50

+0.57 +0.56 +0.67 +0.56 +0.58 +0.60 +0.60 +0.60 +0.54 -0.60 -0.56

+0.67 +0.71 +0.73 +0.64 +0.67 +0.70 +0.67 +0.70 +0.62 +0.67 +0.56 +0.67

-0.62 + 0.54 + 0.53 + 0.59 + 0.53 + 0.56 + 0.56 + 0.56 + 0.56 + 0.53 - 0.54 - 0.53 - 0.50 - 0.62

 $-0.50 \quad +0.57 \quad -0.50 \quad +0.83 \quad +0.70 \quad +0.88 \quad +0.83 \quad +0.73 \quad +0.78 \quad +0.86 \quad +0.86 \quad +0.70 \quad +0.62 \quad +0.60 \quad +0.67$ 

 $-0.50 \quad +0.56 \quad -0.50 \quad +0.70 \quad +0.75 \quad +0.88 \quad +0.75 \quad +0.73 \quad +0.86 \quad +0.78 \quad +0.78 \quad +0.70 \quad +0.62 \quad +0.62 \quad +0.71 \quad -0.50$ 

semantic\_random1 +

semantic\_random2 -

thematic\_random1 -

thematic\_random2 -

thematic\_random3

thematic random5

thematic top to bottom

thematic bottom to top-

deepseek solutions

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

Prompt B

Sign Test Results Between Prompts

+0.71 +0.86 +0.80 +0.75 +0.75 +0.83 +0.83 +0.71 +0.57 +0.57 +0.62 -0.57

+0.57 +0.57 +0.60 +0.60 -0.50 -0.67

-0.57 -0.67 -0.67 -0.60 -0.75 -0.70 -0.83 -0.67 -0.88

-0.60

-0.67 -0.71 -0.60

+0.60 +0.67 +1.00 +0.67 -0.50 -0.62 -0.75 -0.56

 $+0.83 \quad +0.75 \quad +0.88 \quad +0.83 \quad +0.73 \quad +0.78 \quad +0.86 \quad +0.86 \quad +0.70 \quad +0.75 \quad +0.67 \quad +0.71 \quad -0.50$ 

+0.67 -0.50 +0.60 +0.60 +0.60 +0.60 -0.50 -0.67

+0.60 -0.50 +0.75 +0.75 +0.80 +0.75 +0.69 +0.73 +0.78 +0.78 +0.67 +0.75 +0.67 +0.71 -0.50 -0.50 +0.57 -0.50 +0.60

p > 0.05