Sign Test Results Between Prompts

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

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