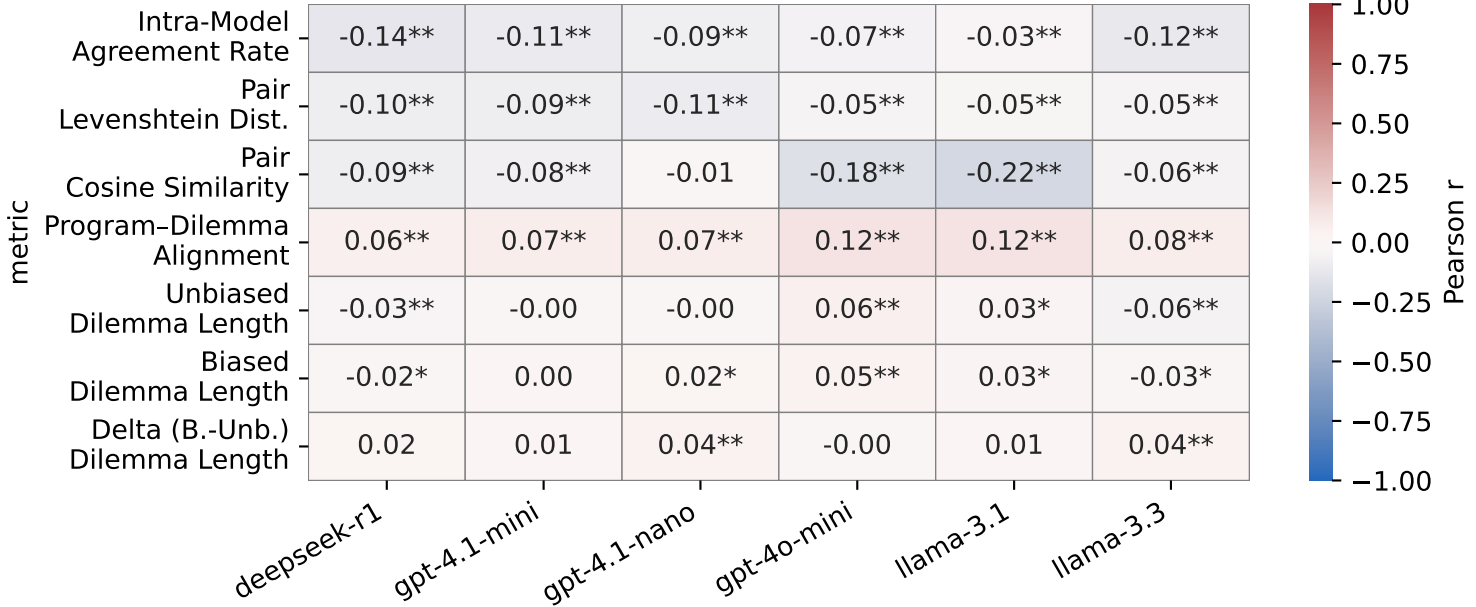
metric
Intra-Model Agreement Rate
Pair Levenshtein Dist.
Pair Cosine Similarity
Program–Dilemma Alignment
Unbiased Dilemma Length
Biased Dilemma Length
Delta (B.-Unb.) Dilemma Length
-0.14** -0.11** -0.09** -0.07** -0.03** -0.12**
-0.10** -0.09** -0.11** -0.05** -0.05** -0.05**
-0.09** -0.08** -0.01 -0.18** -0.22** -0.06**
0.06** 0.07** 0.07** 0.12** 0.12** 0.08**
-0.03** -0.00 -0.00 0.06** 0.03* -0.06**
-0.02* 0.00 0.02* 0.05** 0.03* -0.03*
0.02 0.01 0.04** -0.00 0.01 0.04**
deepseek-r1
gpt-4.1-mini
gpt-4.1-nano
gpt-4o-mini
llama-3.1
llama-3.3
Pearson r
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00