

Performance on simple inferences vs. LMSYS Elo score

Temperature: 0, Condition: chain_of_thought

