

Performance on simple inferences vs. MMLU score

Temperature: 0, Condition: chain_of_thought

