

Performance on simple inferences vs. MMLU score

Temperature: 1, Condition: few-shot

