

Performance on simple inferences vs. GSM-8K score

Temperature: None, Condition: zero-shot_log_probs

