Performance on simple inferences vs. MMLU score
Temperature: None, Condition: zero-shot_log_probs