

Q&A Accuracy vs AUROC: no_abst_raw_logits_second_prompt, mmlu ($r = 0.8$)

