

Q&A Accuracy vs AUROC: no_abst_raw_logits_first_prompt, mmlu ($r = 0.84$)

