

Q&A Accuracy vs AUROC: no_abst_raw_logits_first_prompt, hellaswag ($r = 0.7$)

