

Q&A Accuracy vs AUROC: yes_abst_raw_logits_first_prompt, hellaswag ($r = 0.9$)

