

Q&A Accuracy vs AUROC: yes_abst_raw_logits_second_prompt, hellaswag (r = 0.96)

