

Q&A Accuracy vs AUROC: no_abst_raw_logits_second_prompt, hellaswag (r = 0.92)

