

Q&A Accuracy vs AUROC: yes_abst_raw_logits, hellaswag ($r = 0.76$)

