

Q&A Accuracy vs AUROC: yes\_abst\_raw\_logits\_first\_prompt, hellaswag (r = 0.9)

