

Q&A Accuracy vs AUROC: no\_abst\_raw\_logits\_second\_prompt, mmlu ( $r = 0.8$ )

