

Q&A Accuracy vs AUROC: no\_abst\_raw\_logits\_first\_prompt, hellaswag ( $r = 0.7$ )

