

Score vs Confidence Threshold: no\_abst\_raw\_logits\_second\_prompt, hellaswa

