

Score vs Confidence Threshold: yes\_abst\_raw\_logits\_first\_prompt, winogrande

