

Score vs Confidence Threshold: yes abst raw logits first_prompt, truthfulqa

