

core vs Confidence Threshold: yes\_abst\_raw\_logits\_second\_prompt, winogran

