

(harsh) vs Confidence Threshold: no\_abst\_norm\_logits\_second\_prompt, wino

