

### Score vs Confidence Threshold: no abst raw logits first\_prompt, hellaswag

