Score (Conservative) vs Confidence Threshold: Max Logit, truthfulga 25 Falcon 40B Falcon 7B 20 Llama2 13B Llama2 70B 15 Llama2 7B Score (Conservative) Mistral 7B 10 Mixtral 8x7B SOLAR 10.7B Yi 34B 5 Yi 6B 0 **-**5 -10-1510 20 30 40 50

Confidence Threshold