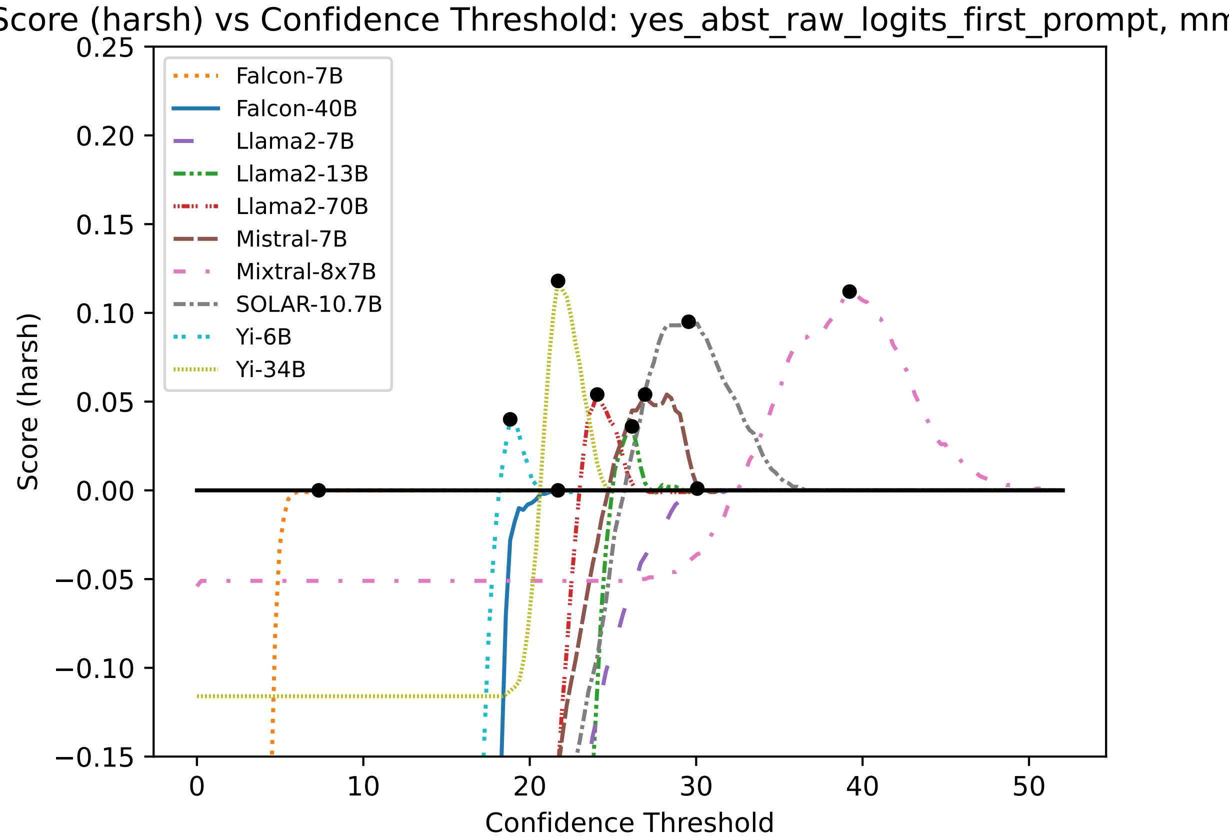
Score (harsh) vs Confidence Threshold: yes_abst_raw_logits_first_prompt, mm
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
Score (harsh)
0
10
20
30
40
50
Confidence Threshold
Falcon-7B
Falcon-40B
Llama2-7B
Llama2-13B
Llama2-70B
Mistral-7B
Mixtral-8x7B
SOLAR-10.7B
Yi-6B
Yi-34B