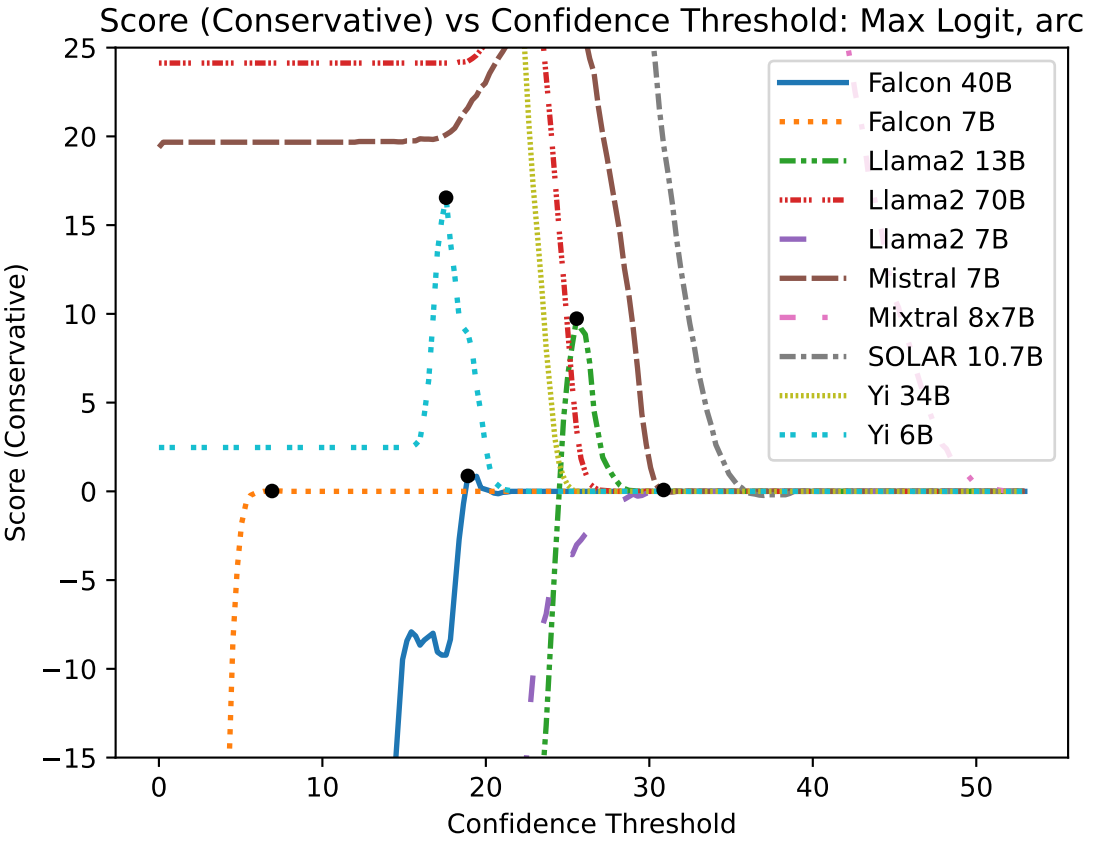
Score (Conservative) vs Confidence Threshold: Max Logit, arc
Score (Conservative)
Confidence Threshold
25
20
15
10
5
0
−5
−10
−15
0
10
20
30
40
50
Falcon 40B
Falcon 7B
Llama2 13B
Llama2 70B
Llama2 7B
Mistral 7B
Mixtral 8x7B
SOLAR 10.7B
Yi 34B
Yi 6B