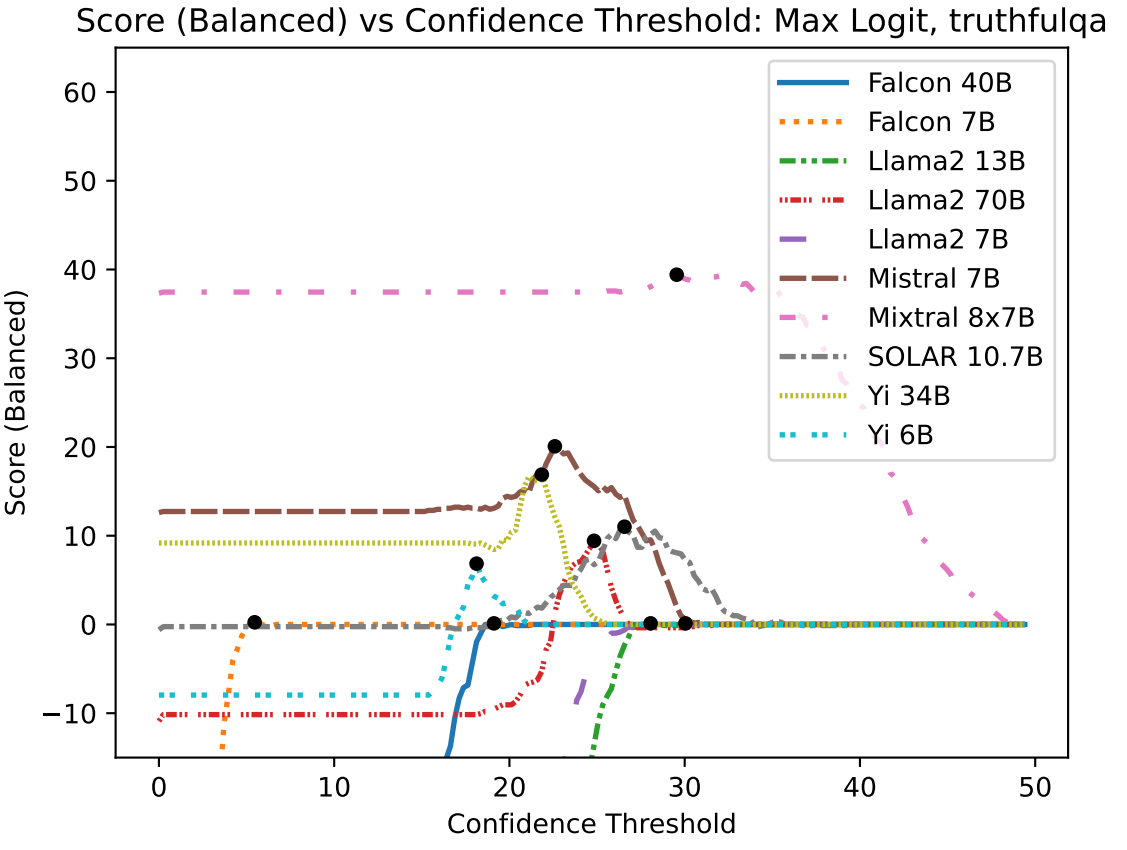
Score (Balanced) vs Confidence Threshold: Max Logit, truthfulqa
Score (Balanced)
Confidence Threshold
60
50
40
30
20
10
0
−10
0
10
20
30
40
50
Falcon 40B
Falcon 7B
Llama2 13B
Llama2 70B
Llama2 7B
Mistral 7B
Mixtral 8x7B
SOLAR 10.7B
Yi 34B
Yi 6B