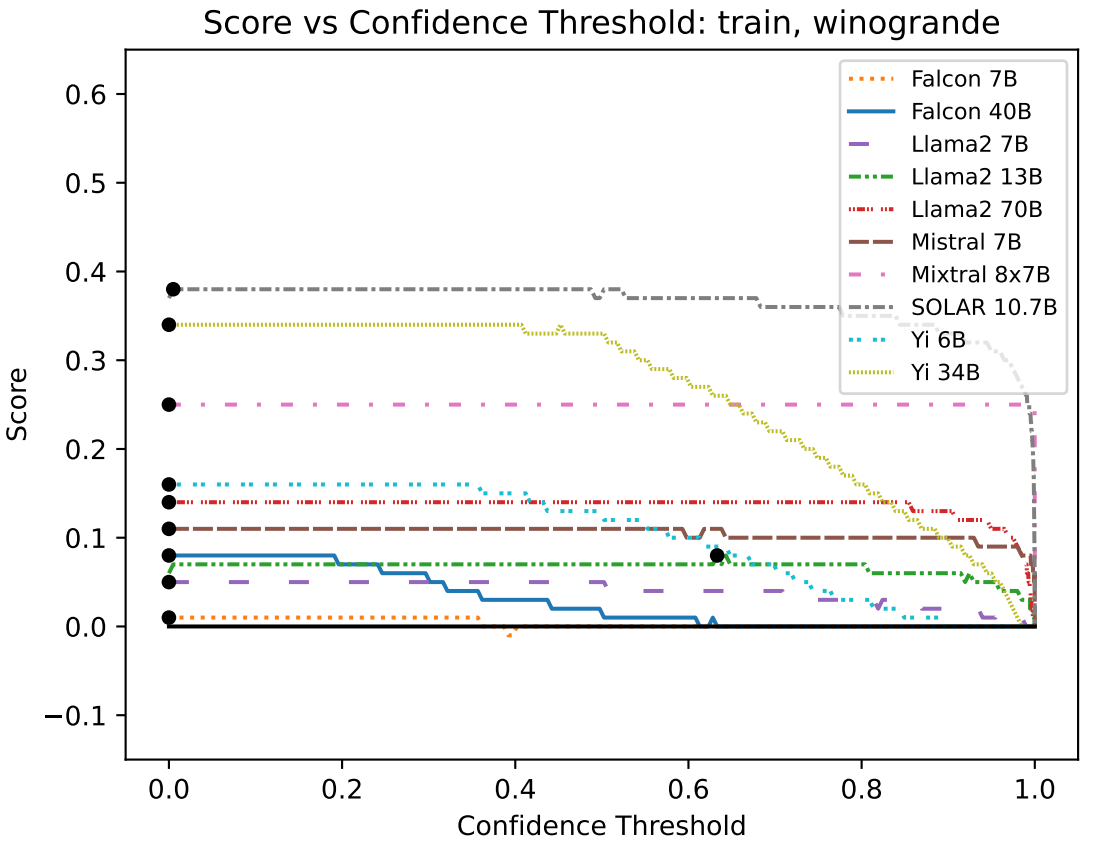
Score vs Confidence Threshold: train, winogrande
Falcon 7B
Falcon 40B
Llama2 7B
Llama2 13B
Llama2 70B
Mistral 7B
Mixtral 8x7B
SOLAR 10.7B
Yi 6B
Yi 34B
Score
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
0.0
0.2
0.4
0.6
0.8
1.0
Confidence Threshold