

Score (Balanced) vs Confidence Threshold: Max Logit, mmlu

