

### Score vs Confidence Threshold: test, mmlu

