

MMLU, phi-4-mini-reasoning to GPT-4.1 mini

Accuracy

0.850
0.825
0.800
0.775
0.750
0.725
0.700

0.00

0.25

0.50

0.75

1.00

Routing Ratio

- average-token-prob
- verbalization-1s
- verbalization-2s
- p(true)
- trained-probe
- perplexity
- jaccard-degree
- ood-probe