

PIQA, Qwen3-4B to DeepSeek-R1

Accuracy

0.96
0.94
0.92
0.90
0.88
0.86

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