

# AQuA, Qwen3-0.6B to Llama-3.1-70B



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