

PubMedQA

 $p_{\theta}^{\text{Choices}}(\text{Correct Choice})$ Per-Sample Correlation(Compute, Score) for
Pythia 300B Tokens (Param. Scaling)