

HellaSwag

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