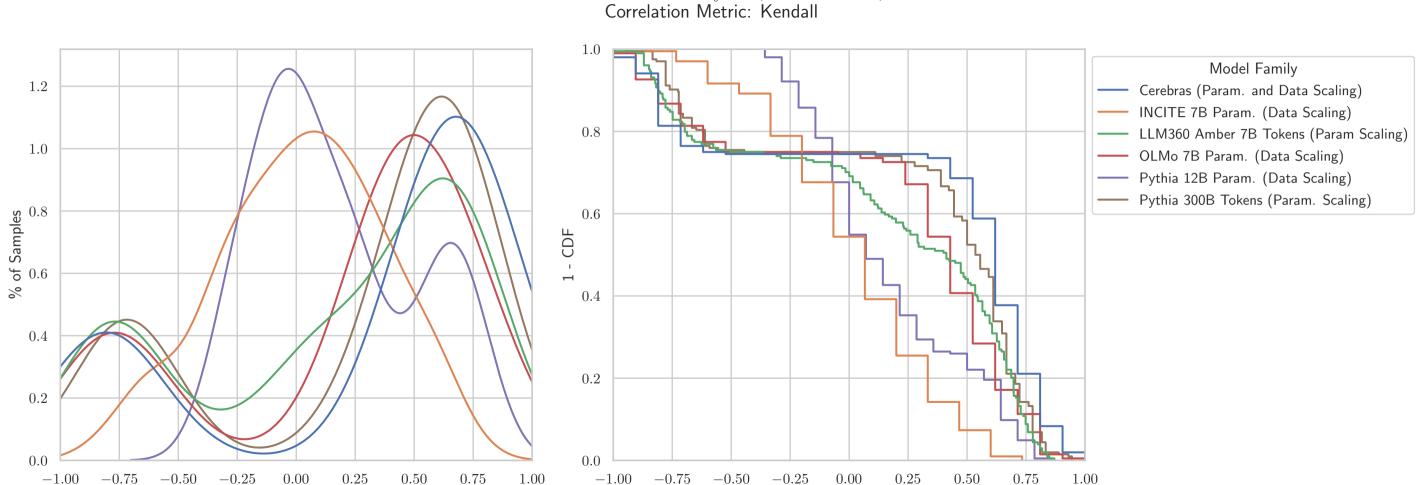
Benchmark: MMLU HS US History Performance Metric: $p_{\theta}^{\text{Choices}}(\text{Correct Choice})$



Correlation Between FLOPs and Scores (Per Sample)

Correlation Between FLOPs and Scores (Per Sample)