

Distributions of Score-Compute Correlations by Metric
Benchmark: HellaSwag

