

Benchmark: MMLU Moral Disputes  
Performance Metric: Brier Score  
Correlation Metric: Spearman

