

Benchmark: MMLU Human Aging  
Performance Metric: Brier Score  
Correlation Metric: Kendall

