

Benchmark: MMLU Human Aging  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Pearson

