

Benchmark: MMLUHS Psychology  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Spearman

