

Benchmark: MMLU Professional Psychology
Performance Metric: Target Greedily Decoded
Correlation Metric: Pearson

