Benchmark: MMLU Abstract Algebra Performance Metric: $p_{\theta}^{\text{Choices}}(\text{Correct Choice})$ Correlation Metric: Spearman

