gpt4.1-mini - Phase 1: Comprehensive Metrics
(2 experiments average)

	Accuracy	Macro-F1	Weighted-F1
Type	0.633	0.670	0.630
Party	0.750	0.804	0.750
Action	0.483	0.282	0.462
Asset	0.758	0.228	0.690