

Model Performance Comparison: Standard vs Rich Prompts

Model	Approach	Accuracy (Standard)	Accuracy (Rich)	ΔAccuracy	Precision (Standard)	Precision (Rich)	Δ Precision	F1-Score (Standard)	F1-Score (Rich)	Δ F1-Score
claude-sonnet	few_shot → few_sho	54.520	80.000	25.480	58.540	87.500	28.960	41.380	77.780	36.400
gemini	few_shot → few_sho	52.510	84.330	31.820	55.260	81.210	25.950	37.170	85.080	47.910
gpt-3.5-turbo	few_shot → few_sho	62.880	61.000	-1.880	76.710	61.870	-14.840	50.220	59.520	9.300
gpt-4o	few_shot → few_sho	74.580	66.000	-8.580	80.830	98.000	17.170	71.850	49.000	-22.850
gpt-4o-mini	few_shot → few_sho	58.190	65.670	7.480	79.070	82.190	3.120	35.230	53.810	18.580