Comparative Analysis of Model Metrics (Publication Optimized)

	accuracy	precision	recall	f1	true_positives	false_positives	true_negatives	false_negatives
Original Label	0.50	0.09	0.09	0.48	42.00	19.00	68.00	8.00
DEA_found_index_terms	0.63	1.00	0.38	0.51	45.00	11.00	56.00	11.00
pubchem_found_index_terms	0.37	0.38	0.76	0.79	45.00	26.00	52.00	19.00
GPT_found_index_terms	0.15	0.85	0.25	0.42	29.00	23.00	61.00	16.00
redmed_found_index_terms	0.07	0.41	0.74	0.18	41.00	13.00	54.00	17.00
found_index_terms	1.00	0.31	0.63	0.92	17.00	20.00	50.00	19.00
o4mini_label_short	0.06	0.61	0.17	0.52	34.00	12.00	48.00	7.00
4o_mini_label_short	0.57	0.10	0.14	0.73	27.00	27.00	42.00	18.00
llama_label_short	0.61	0.98	0.91	0.21	31.00	10.00	49.00	17.00
qwen_4b_label_short	0.89	0.16	0.36	0.96	20.00	16.00	61.00	9.00
o4mini_label	0.49	0.98	0.70	0.88	42.00	11.00	66.00	15.00
4o_mini_label	0.30	0.83	0.73	0.21	12.00	17.00	53.00	8.00
llama_label	0.31	0.12	0.45	0.23	12.00	25.00	67.00	17.00
qwen_4b_label	0.19	0.43	0.37	0.57	29.00	20.00	56.00	12.00
o4mini_drug_labels	0.99	0.34	0.31	0.78	32.00	17.00	68.00	7.00
4o_mini_drug_labels	0.42	0.42	0.36	0.90	31.00	9.00	57.00	8.00
llama_drug_labels	0.50	0.93	0.02	0.04	31.00	8.00	50.00	9.00
qwen_4b_drug_labels	0.58	0.36	0.35	0.79	19.00	27.00	44.00	19.00
o4mini_rag	0.37	0.60	0.57	0.77	24.00	22.00	69.00	8.00
4o_mini_rag	0.14	0.95	0.59	0.95	26.00	18.00	66.00	15.00
llama_rag	0.90	0.58	0.46	0.89	41.00	12.00	48.00	16.00
qwen_4b_rag	0.96	0.85	0.05	0.88	34.00	9.00	55.00	5.00

Normalized Score (Darker = Better Performance)