Model Performance Matrix

Models	Prompt 2 (P2): BETO	Prompt 2 (P2): SciBETO	Prompt 2 (P2): MPNet-Mult	Prompt 2 (P2): XLM-R	Prompt 2 (P2): SciBETO-mean	Prompt 2 (P2): chrF
sonar	★ **0.504**	0.723	★ **0.503**	★ **0.466**	0.678	★ **0.306**
gpt-5.1	0.477	0.735	0.395	0.370	★ **0.709**	0.285
claude-sonnet-4	0.482	★ **0.746**	0.405	0.372	0.694	0.276
gpt-4.1	0.475	0.742	0.399	0.372	0.698	0.275
llama-4-maverick	0.458	0.737	0.368	0.340	0.658	0.247
command-r-08-2024	0.429	0.697	0.330	0.311	0.640	0.227
deepseek-chat-v3.1	0.461	0.733	0.346	0.309	0.689	0.261
mistral-medium-3	0.434	0.708	0.326	0.298	0.645	0.236
gemma-2-27b-it	0.435	0.724	0.312	0.291	0.632	0.210
gemini-2.5-flash	0.465	0.713	0.404	0.380	0.674	0.265
o1-mini	0.456	0.741	0.366	0.332	0.678	0.247
llama-3.3-70b-instruct	0.422	0.702	0.345	0.308	0.583	0.194
nova-lite-v1	0.433	0.733	0.336	0.303	0.619	0.202
llama-3.1-405b-instruct	0.403	0.658	0.303	0.275	0.590	0.213
o4-mini-high	0.454	0.740	0.363	0.329	0.677	0.246
phi-4	0.407	0.687	0.266	0.238	0.632	0.220
mixtral-8x7b-instruct	0.442	0.708	0.343	0.321	0.652	0.235
grok-3-mini-beta	0.439	0.714	0.351	0.321	0.642	0.228
nova-micro-v1	0.441	0.733	0.328	0.299	0.659	0.228
qwen-2.5-72b-instruct	0.407	0.662	0.329	0.301	0.586	0.209
kimi-k2-0905	0.260	0.418	0.202	0.176	0.383	0.138
wizardlm-2-8x22b	0.446	0.717	0.322	0.282	0.675	0.252
qwen2.5-vl-32b-instruct:free						