Models	Task 2: BERTScore BETO	Task 2: BERTScore SciBETO	Task 2: Sentence-BERT MPNet-Mult	Task 2: Sentence-BERT XLM-R	Task 2: Sentence-BERT SciBETO-mean	Task 2: ChrF	Task 3: BERTScore BETO	Task 3: BERTScore SciBETO	Task 3: Sentence-BERT MPNet-Mult	Task 3: Sentence-BERT XLM-R	Task 3: Sentence-BERT SciBETO-mean	Task 3: ChrF
sonar	★ **0.504**	0.723	★ **0.503**	★ **0.466**	0.678	★ **0.306**	★ **0.549**	★ **0.769**	★ **0.651**	★ **0.618**	0.707	0.286
gpt-5.1	0.477	0.735	0.395	0.370	★ **0.709**	0.285	0.544	0.767	0.646	0.604	★ **0.715**	0.277
claude-sonnet-4	0.482	★ **0.746**	0.405	0.372	0.694	0.276	0.531	0.740	0.621	0.591	0.702	★ **0.290**
gpt-4.1	0.475	0.742	0.399	0.372	0.698	0.275	0.528	0.755	0.637	0.599	0.666	0.256
llama-4-maverick	0.458	0.737	0.368	0.340	0.658	0.247	0.508	0.748	0.588	0.552	0.642	0.240
command-r-08-2024	0.429	0.697	0.330	0.311	0.640	0.227	0.510	0.755	0.543	0.503	0.703	0.250
deepseek-chat-v3.1	0.461	0.733	0.346	0.309	0.689	0.261	0.500	0.713	0.572	0.543	0.665	0.264
mistral-medium-3	0.434	0.708	0.326	0.298	0.645	0.236	0.518	0.753	0.586	0.541	0.694	0.253
gemma-2-27b-it	0.435	0.724	0.312	0.291	0.632	0.210	0.475	0.736	0.521	0.480	0.621	0.201
gemini-2.5-flash	0.465	0.713	0.404	0.380	0.674	0.265	0.507	0.731	0.581	0.543	0.663	0.237
o1-mini	0.456	0.741	0.366	0.332	0.678	0.247	0.511	0.752	0.605	0.574	0.656	0.243
llama-3.3-70b-instruct	0.422	0.702	0.345	0.308	0.583	0.194	0.456	0.720	0.527	0.495	0.573	0.187
nova-lite-v1	0.433	0.733	0.336	0.303	0.619	0.202	0.503	0.759	0.539	0.494	0.688	0.240
llama-3.1-405b-instruct	0.403	0.658	0.303	0.275	0.590	0.213	0.464	0.687	0.520	0.485	0.614	0.227
o4-mini-high	0.454	0.740	0.363	0.329	0.677	0.246	0.494	0.727	0.583	0.552	0.636	0.234
phi-4	0.407	0.687	0.266	0.238	0.632	0.220	0.474	0.734	0.469	0.431	0.682	0.235
mixtral-8x7b-instruct	0.442	0.708	0.343	0.321	0.652	0.235	0.451	0.677	0.498	0.464	0.620	0.229
grok-3-mini-beta	0.439	0.714	0.351	0.321	0.642	0.228	0.472	0.707	0.559	0.532	0.617	0.224
nova-micro-v1	0.441	0.733	0.328	0.299	0.659	0.228	0.447	0.723	0.509	0.474	0.578	0.183
qwen-2.5-72b-instruct	0.407	0.662	0.329	0.301	0.586	0.209	0.443	0.668	0.501	0.468	0.574	0.195
kimi-k2-0905	0.260	0.418	0.202	0.176	0.383	0.138	0.453	0.657	0.530	0.497	0.590	0.216
wizardlm-2-8x22b	0.446	0.717	0.322	0.282	0.675	0.252						
qwen2.5-vl-32b-instruct:free												

Metrics

Best Model

Score

Worst Model