

Harmonic Score per Model and Prompt

Model

LLaMA 3.1 8B	0.462	0.438	0.442	0.751
Zephyr 7B	0.549	0.582	0.503	0.400
OpenChat 3.5	0.538	0.541	0.469	0.413
OpenChat 3.5-0106	0.424	0.442	0.438	0.528
InternLM2.5 7B	0.517	0.455	0.387	
OpenChat 3.5 GPTQ	0.477	0.497	0.410	0.402
Qwen2 7B	0.539	0.458	0.343	0.412
Nous Hermes Mistral 7B DPO				0.425
Nous Hermes Yi 34B	0.449	0.451	0.388	0.379
LLaMA 2 13B				0.384
LLaMA 3 8B				0.367
Yi 6B	0.240	0.337	0.361	
Gemma 7B	0.397	0.418	0.007	0.404
LLaMA 2 7B	0.000	0.000	0.000	0.385
GPT-2-Small	0.180	0.000	0.000	
StableBeluga 7B	0.000	0.000	0.004	

Prompt

prompt1

prompt2

prompt3

prompt4

0.7

0.6

0.5

0.4

0.3

0.2

0.1

0.0