Performance-Quality Trade-off Matrix (Lower is Better)

model_display	prompt1	prompt2	prompt3	prompt4
GPT-2-Large	0.18	0.53	0.18	
GPT-2-Medium	0.19	0.38	0.23	
GPT-2-Small	0.27	0.28	0.28	
Gemma 7B	0.26	0.22	0.09	0.16
InternLM2.5 7B	0.22	0.30	0.34	0.64
LLaMA 3.1 8B	0.08	0.25	0.55	0.29
Mistral 7B	0.12	0.09	0.13	0.03
Nous Hermes Yi 34B	0.05	0.08	0.16	0.03
OpenChat 3.5	0.15	0.15	0.25	0.40
OpenChat 3.5 GPTQ	0.10	0.15	0.20	0.62
Qwen2 7B	0.02	0.07	0.21	0.07
StableBeluga 7B	0.11	0.17	0.17	0.24
Yi 6B	0.21	0.03	0.39	0.38
Zephyr 7B	0.09	0.06	0.13	0.09

prompt