

Harmonic Privacy-Quality Score by Model and Prompt

Model

GPT-2-Large	0.403	0.380	0.268	
GPT-2-Medium	0.312	0.312	0.459	
GPT-2-Small	0.330	0.312	0.285	
Gemma 7B	0.400	0.602	0.287	0.548
InternLM2.5 7B	0.500	0.653	0.632	0.495
LLaMA 2 13B	0.327	0.312	0.554	0.368
LLaMA 2 13B Chat	0.396	0.390	0.511	0.682
LLaMA 2 7B	0.312	0.312	0.421	0.398
LLaMA 3 8B	0.312	0.397	0.533	0.546
LLaMA 3.1 8B	0.506	0.618	0.553	0.599
Mistral 7B	0.588	0.565	0.720	0.574
Nous Hermes Mistral 7B DPO	0.569	0.534	0.603	0.573
Nous Hermes Yi 34B	0.601	0.611	0.568	0.588
OpenChat 3.5	0.636	0.715	0.696	0.679
OpenChat 3.5 GPTQ	0.611	0.655	0.627	0.574
OpenChat 3.5-0106	0.567	0.659	0.680	0.613
Qwen2 7B	0.485	0.484	0.454	0.538
StableBeluga 7B	0.312	0.312	0.555	0.512
Yi 6B	0.740	0.627	0.501	0.679
Zephyr 7B	0.550	0.498	0.660	0.546
	prompt1	prompt2	prompt3	prompt4

Prompt

0.3

0.4

0.5

0.6

0.7