

Model	Medical Accuracy	Completeness	Safety	Clarity	Helpfulness	Overall
Llama-3.3-70B	4.83	4.46	3.96	4.53	4.09	4.4
Llama-4-Maverick	4.32	4.26	3.83	4.97	4.18	4.23
Llama-3-8B	4.21	3.7	3.34	3.93	3.59	3.77
Llama-3.2-3B	3.53	2.95	3.05	3.43	2.94	3.2
GPT-5-Mini	3.1	2.9	3.8	3.2	2.8	3.16