

Source	N	Accuracy (0-2)	Completeness (0-2)	Empathy (0-2)	Safety (0-1)	Hallucinations (0-1)	Acceptability (0-5)*
LLM_A (ChatGPT)	72	1.9	1.7	1.2	1.0	0.01	4.6
LLM_B (Gemini)	69	1.9	1.8	1.3	0.9	0.01	4.6
LLM_C (Google)	76	1.8	1.7	0.2	1.0	0.01	4.5
Physician	72	1.4	1.0	0.04	0.9	0.01	3.3