Mean Diagnostic Score by Model and Clinical Domain

g

Model	Neurology	Pulmonology	Psychiatry	ENT
GPT	4.12	4.40	3.94	4.83
DeepSeek	4.10	4.33	3.78	4.58
Claude	3.88	4.41	3.50	5.00
Gemini	3.84	4.26	3.78	4.17
Grok	3.88	4.14	3.48	4.83