Model performance by domain

Domain	n
Ent	n=5
Psychiatry	n=16
Pulmonology	n=30
Neurology	n=74

Mean score

Grok · Claude · DeepSeek · GPT · Gemini