Model Accuracy Across Knowledge Groups

Overall r = 0.75 | JT p = 2.54e-07

Anthropic claude-sonnet-4 (r=0.97)
xAI grok-3-mini (r=0.86)

DeepSeek deepseek-chat-v3-0324 (r=0.97)
OpenAI gpt-5 (r=1.00)
OpenAI gpt-5-nano (r=0.92)
OpenAI gpt-4o (r=1.00)
OpenAI gpt-oss-20b (r=0.74)
OpenAI gpt-4.1-nano (r=0.83)
OpenAI gpt-4 (r=1.00)

Google gemini-2.5-flash-lite (r=0.99)