Diagnostic Capabilities

Argument
Accuracy

- Llama 3.3 70B (Groq)
- Mistral Small
- Mistral Medium
- Llama 4 Scout 17B (Groq)
- Gemini 2.0 Flash (OpenRouter)
- Granite 3B

100%
75%
50%
25%

Data
Flow

Tool
Selection

Completion
Rate