

HLE-Verified Leaderboard

Gemini 3 Pro  48.2%

Claude Opus 4.6  46.8%

GPT-5.2  43.3%

Claude Opus 4.5  38.8%

Qwen3-Max-Thinking  38.2%

DeepSeek-V3.2  36.4%

Grok 4.1 (Fast)  29.0%



Average Accuracy (Pass@1)