

Rogue Deep Think (Preview)

Software engineering benchmarks

DeepSWE v1.1

SCORE (%)

01	Rogue Deep Think (Preview)	75.22%
02	Gemini 3.8 Flash (high)	73.83%
03	GPT-6 Astra (Max)	73.23%
04	GLM 5.3 (Max)	68.96%
05	Kimi K3 (Max)	68.51%
06	Fable 5.1 (Max)	67.4%
07	Grok 4.6 (xhigh)	66.74%
08	DeepSeek V4 Pro (Max)	62.83%
09	Qwen 3.8 Max (xhigh)	57.46%

FrontierSWE v2

SCORE (%)

01	GPT-6 Astra (Max)	65.51%
02	Fable 5.1 (Max) *	56.29%
03	Rogue Deep Think (Preview)	36.78%
04	GLM 5.3 (Max)	30.18%
05	Kimi K3 (Max)	25.87%
06	Grok 4.6 (Max)	25.29%
07	Gemini 3.8 Flash (Max)	19.63%
08	Qwen 3.8 Max (Max)	15.81%

Preview · medium effort. Full power soon.

Base: GLM 5.3 · * Safety fallback

Sources: Datacurve · FrontierSWE · Anthropic