

Rogue Deep Think (Preview)

Measured at medium effort.

DeepSWE v1.1

75.22%

+6.19 pp vs. internal baseline

FrontierSWE v2

36.78%

+6.60 pp vs. internal baseline

Full power. Coming soon.

