
B
Q
GPT-5.6 Sol / Codex CLI
DeepSeek V4.1 Flash / Codex CLI
GPT-5.6 Sol / Pi
DeepSeek V4.1 Flash / Pi
GPT-6.1 Sol / Codex CLI
GPT-6 Astra / Codex CLI
Kimi K3 / Kimi hosted
DeepSeek V4.1 Flash / DSH
Gemini 4 Argon / Gemini CLI
Claude Opus 5.5 / Pi
Claude Opus 5.5 / Claude Code CLI
GPT-5.6 Sol / DSH
Claude Opus 5.5 / Codex CLI
Claude Opus 5.5 / DSH
GPT-6 Luna / Codex CLI
Seed-2.1-pro / Trae Work
DeepSeek V4.1 Flash / WorkBuddy
DuMate / DuMate CLI
60
80
100
Mean score
(a)
<5
5–15
≥15 pts
2.6
13.0
3.0
42.6
1.7
0.4
27.8
29.6
17.4
10.4
23.0
6.5
15.2
21.3
2.6
37.4
31.7
39.6
0
10
20
30
40
50
Responses with Q < B (%)
(b)