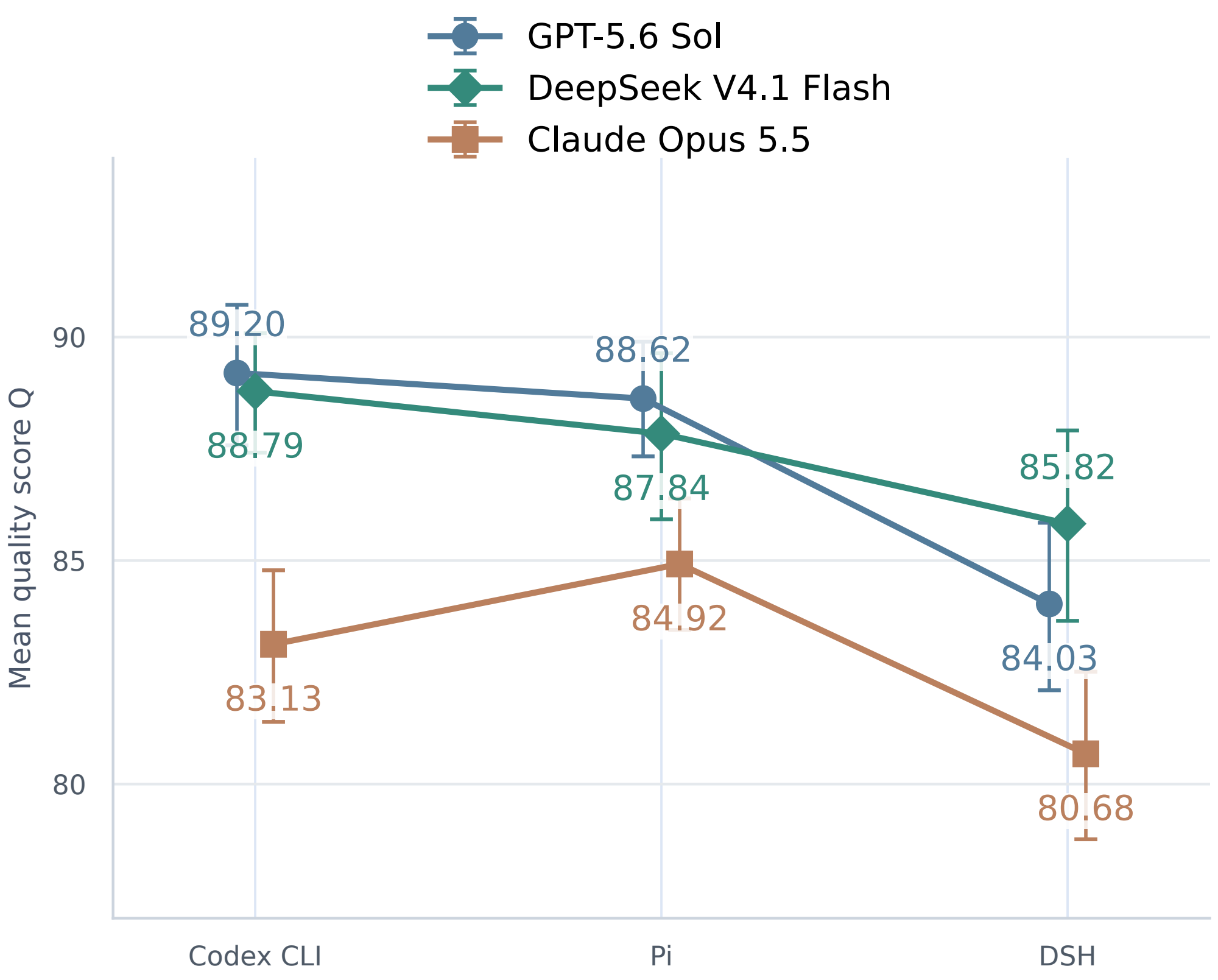

GPT-5.6 Sol
DeepSeek V4.1 Flash
Claude Opus 5.5
Mean quality score Q
90
85
80
89.20
88.79
88.62
87.84
85.82
84.03
84.92
83.13
80.68
Codex CLI
Pi
DSH