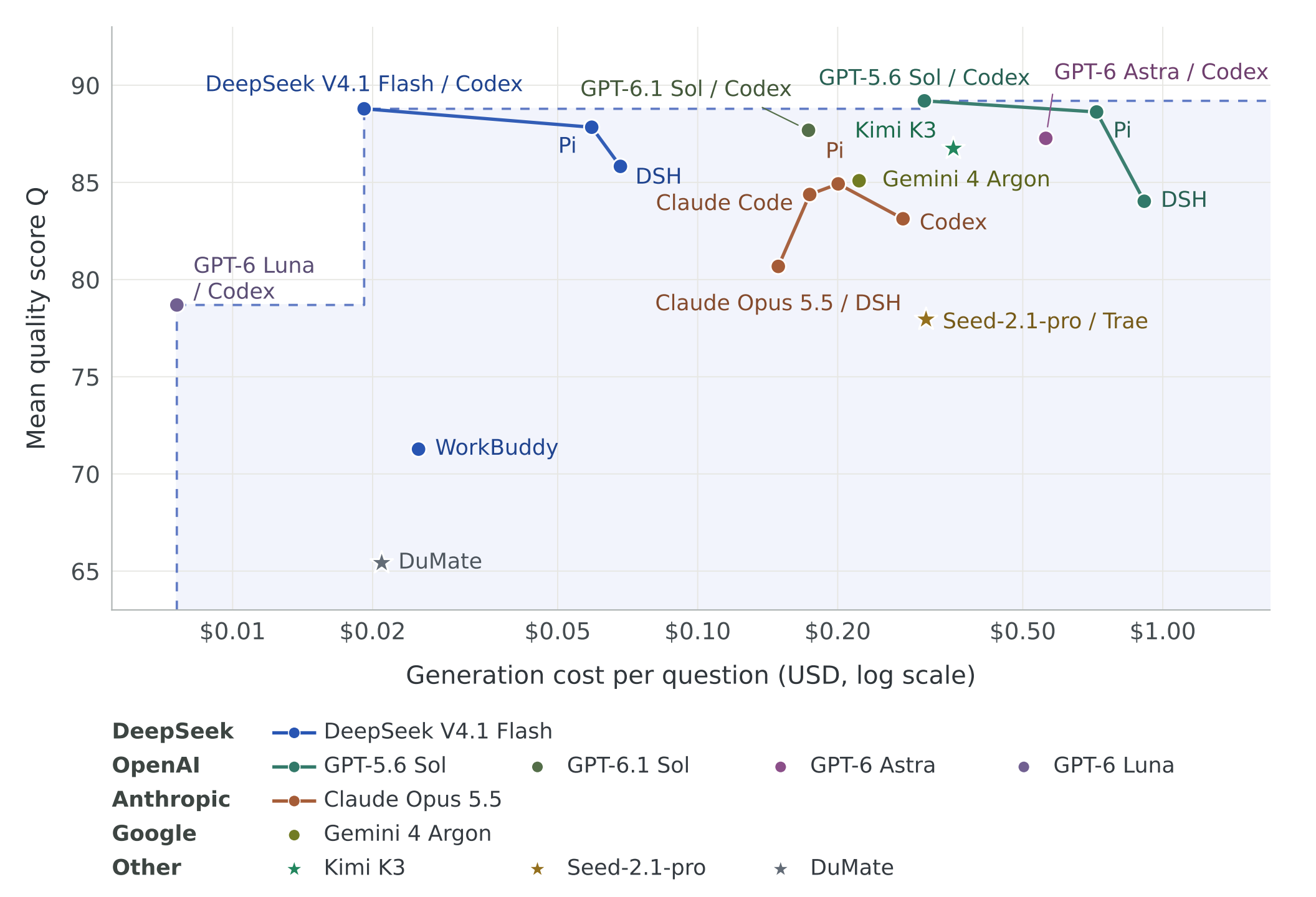

Mean quality score Q
90
85
80
75
70
65
$0.01
$0.02
$0.05
$0.10
$0.20
$0.50
$1.00
Generation cost per question (USD, log scale)
DeepSeek V4.1 Flash / Codex
Pi
DSH
WorkBuddy
DuMate
GPT-6 Luna / Codex
GPT-6.1 Sol / Codex
Pi
Kimi K3
Gemini 4 Argon
Claude Code
Codex
Claude Opus 5.5 / DSH
Seed-2.1-pro / Trae
GPT-5.6 Sol / Codex
GPT-6 Astra / Codex
Pi
DSH
DeepSeek
DeepSeek V4.1 Flash
OpenAI
GPT-5.6 Sol
GPT-6.1 Sol
GPT-6 Astra
GPT-6 Luna
Anthropic
Claude Opus 5.5
Google
Gemini 4 Argon
Other
Kimi K3
Seed-2.1-pro
DuMate