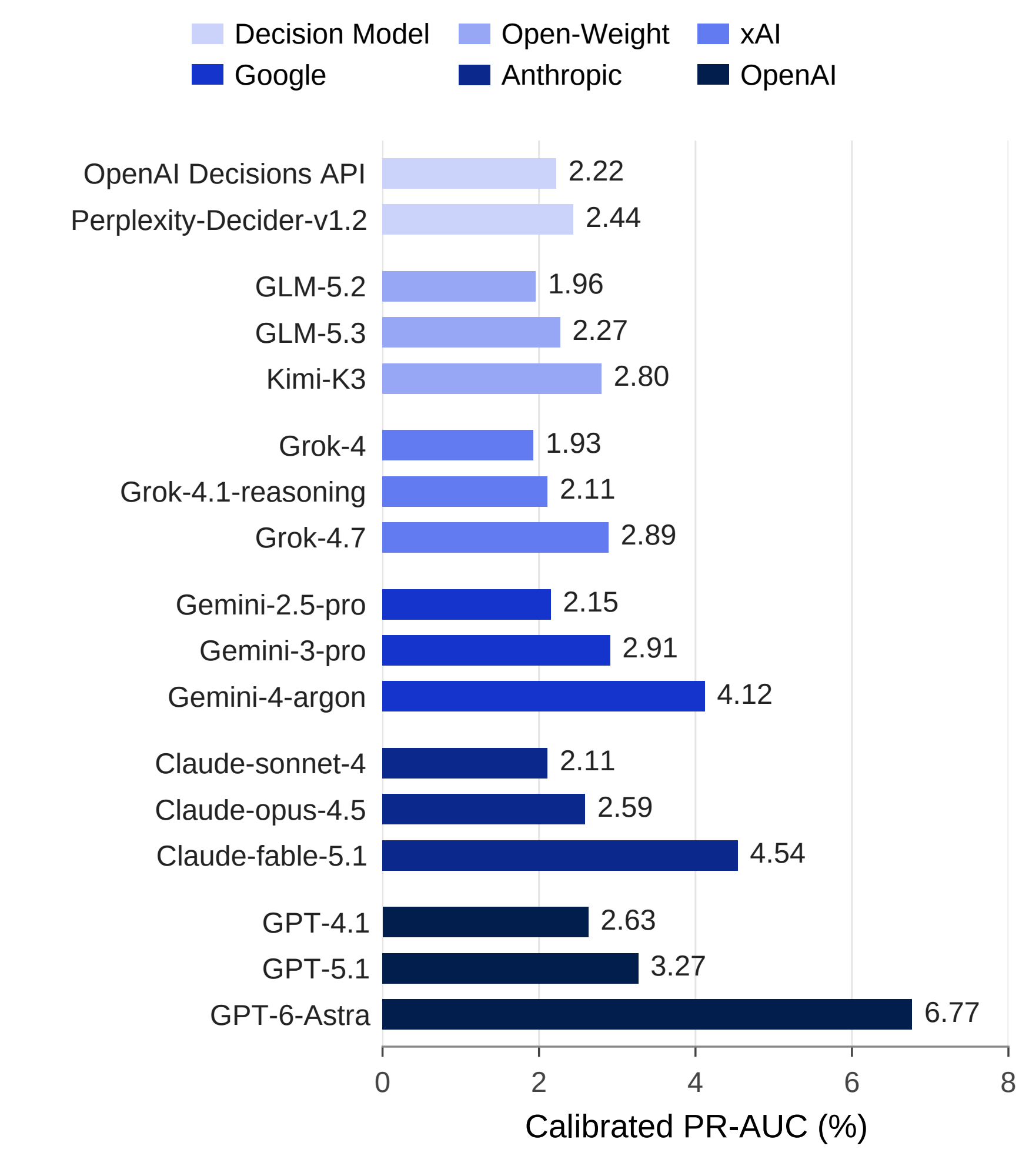

Decision Model
Open-Weight
xAI
Google
Anthropic
OpenAI
OpenAI Decisions API
2.22
Perplexity-Decider-v1.2
2.44
GLM-5.2
1.96
GLM-5.3
2.27
Kimi-K3
2.80
Grok-4
1.93
Grok-4.1-reasoning
2.11
Grok-4.7
2.89
Gemini-2.5-pro
2.15
Gemini-3-pro
2.91
Gemini-4-argon
4.12
Claude-sonnet-4
2.11
Claude-opus-4.5
2.59
Claude-fable-5.1
4.54
GPT-4.1
2.63
GPT-5.1
3.27
GPT-6-Astra
6.77
0
2
4
6
8
Calibrated PR-AUC (%)