First−Try Success Rate on DBCode−Bench
Default model settings

First−Try Success Rate (%) vs **Model**

- Claude Sonnet 4: 33.3%
- Gemini 2.5 Pro: 37.0%
- GPT−4o: 13.6%
- GPT−5: 38.3%
- gpt−oss−120b: 18.5%
- Kimi K2 0905: 23.5%
- o3: 29.6%
- Qwen3 Coder: 21.0%

Model
- Claude Sonnet 4
- Gemini 2.5 Pro
- GPT−4o
- GPT−5
- gpt−oss−120b
- Kimi K2 0905
- o3
- Qwen3 Coder