Empirical Pass@k
Default model settings

Pass@k

Model

- Claude Sonnet 4 (0k Thinking Tokens)
- claude-sonnet-4.5_thinking0
- claude-sonnet-4.5_thinking24k
- claude-sonnet-4.5_thinking8k
- Gemini 2.5 Pro
- GPT-4o
- GPT-5 (Medium Reasoning)
- gpt-oss-120b (Medium Reasoning)
- Kimi K2 0905
- o3
- Qwen3 Coder