
Performance Improvement (%, ↑)
Increased Runtime (%, ↓)
Increased Lines of Code (%, ↓)
i-th implementation in a trial
i-th implementation in a trial
i-th implementation in a trial
claude-3-5-sonnet-v2
gemini-exp-1206
llama3-1-405b-instruct
o3-mini
gpt-4o