

Average Build Success Rate Across Reasoning Configurations

Claude Sonnet 4: 0–24k thinking token budgets on DBCode–Bench

