

Average Build Success Rate Across Reasoning Configurations

GPT-5: Low, Medium, and High Reasoning Efforts on DBCode-Bench

