

Official LongBench metrics (7 tasks, 2,048-token prompts)

