

## Top models per tasks

Best in Reasoning (GPQA Diamond) ⓘ



Best in High School Math (AIME 2025) ⓘ



Best in Agentic Coding (SWE Bench) ⓘ

