

Performance

100%  
80%  
60%  
40%  
20%  
0%

2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030

- Aider Polyglot
- WeirdML
- METR Time Horizons
- SWE-Bench Bash Only
- SWE-Bench Verified
- SWE-Bench Pro
- GSO-Bench

