

Performance

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

2022 2023 2024 2025 2026 2027 2028 2029 2030

- Adversarial NLI
- SimpleBench
- BALROG
- Chess Puzzles
- EnigmaEval

