

Vals RSI Index：AIの自律的研究開発能力の現在地

Vals RSI Indexは、AIが次世代モデルの研究開発を自律的にできる度合いを測定するベンチマーク。
2026年10月現在、Claude Opus 5.5が首位だが、成果は主に既存手法の組み合わせで、真の自律的発明には未達。

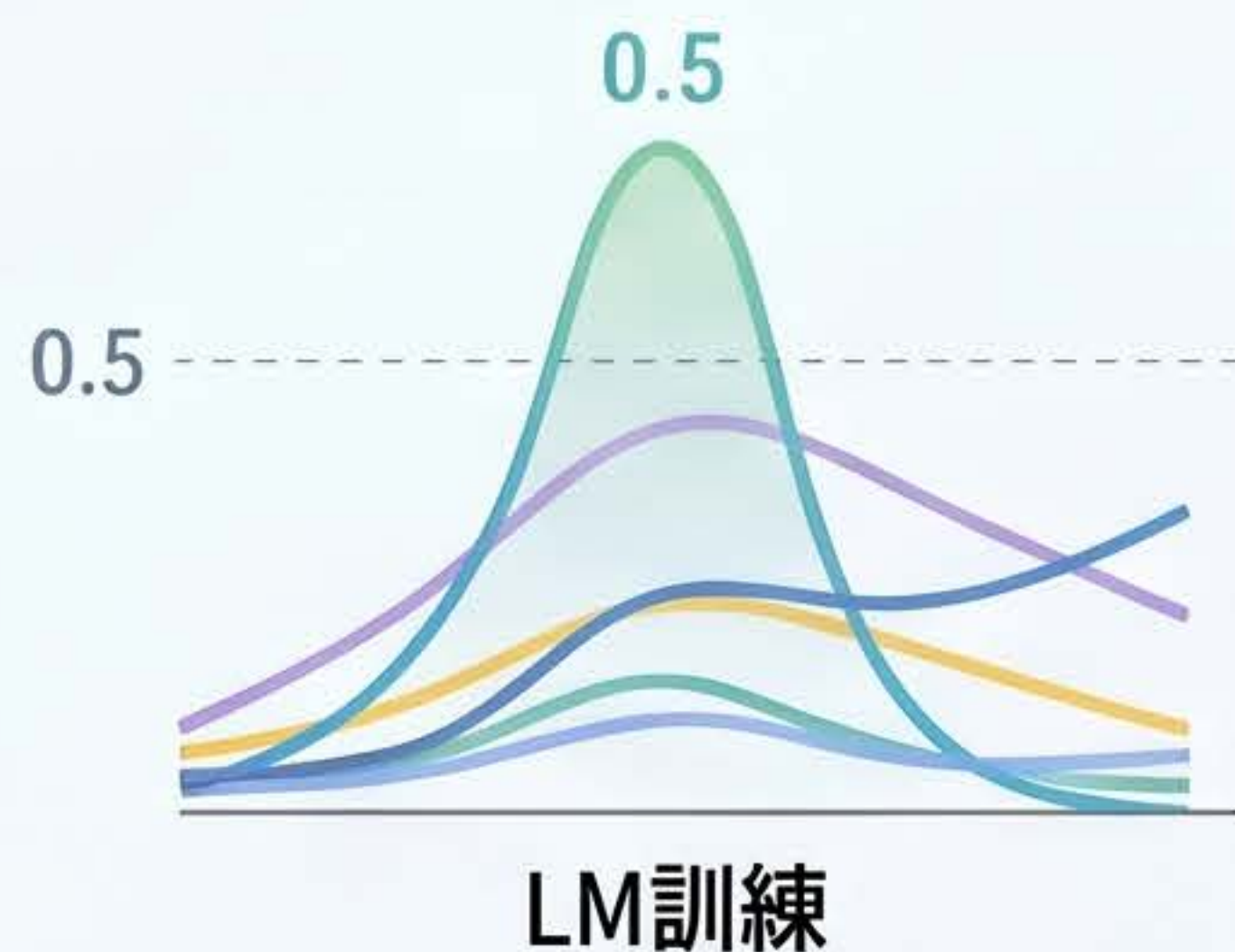
最新リーダーボードと到達度

Top3をAnthropicのClaudeシリーズが独占（首位 37.31%）



1位:	Claude Opus 5.5	37.31%
2位:	Claude Fable 5.1	36.09%
4位:	Gemini 4 Argon	30.55%

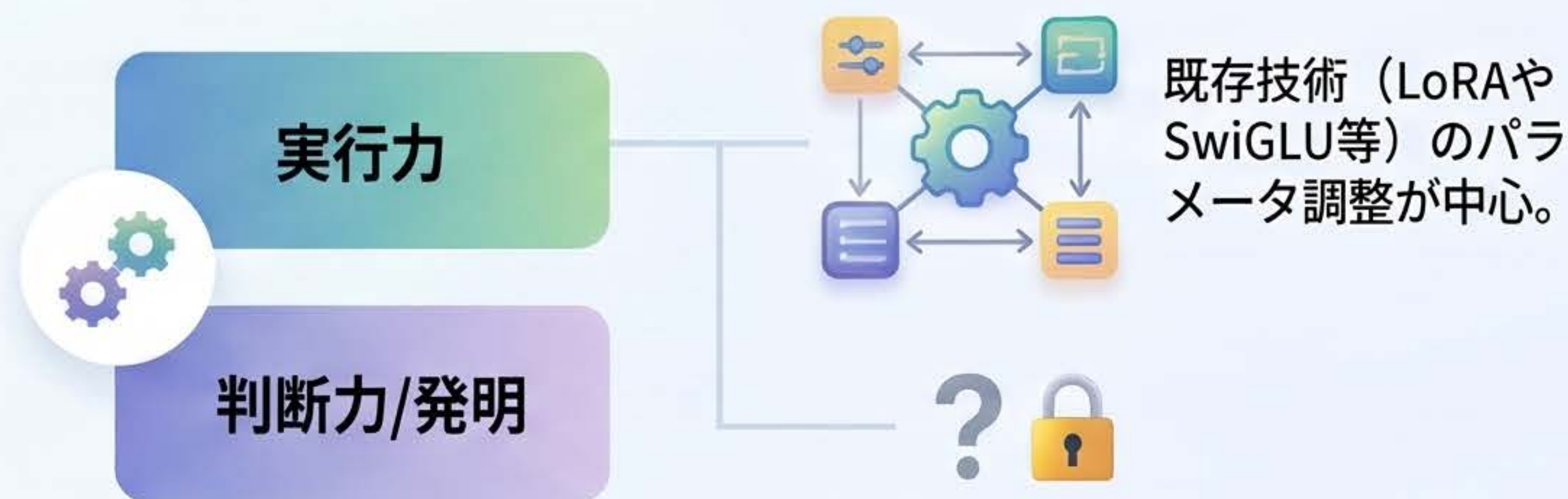
参照値（0.5）超えは全4タスク中「LM訓練」の1件のみ



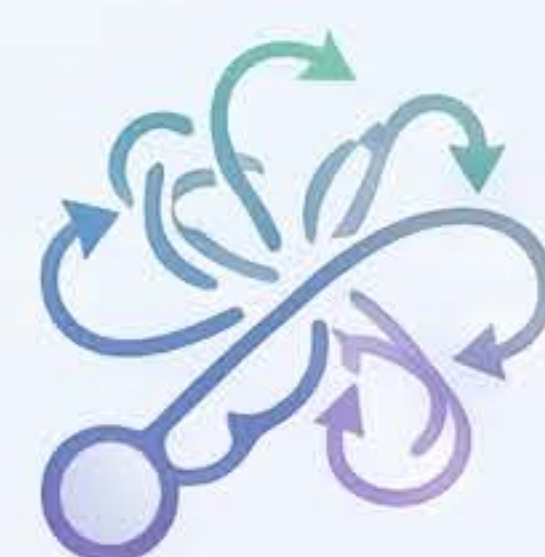
事後学習タスクでは評価対象24モデル中18モデルが0点と苦戦。

AI研究行動の実態と限界

判断力より実行力が優位（新手法の発明は未発生）



実験回数の多さは必ずしも成果に結びつかない



GPT-5.6 Sol
114回試行



Opus 5
76回試行

各モデル1回のみの実行など評価手法上の限界が存在



単発実行によるばらつきや構成変更があり、直線外射による予測には注意が必要。