

Vals RSI Indexの真実：AI自律研究開発能力の現在地

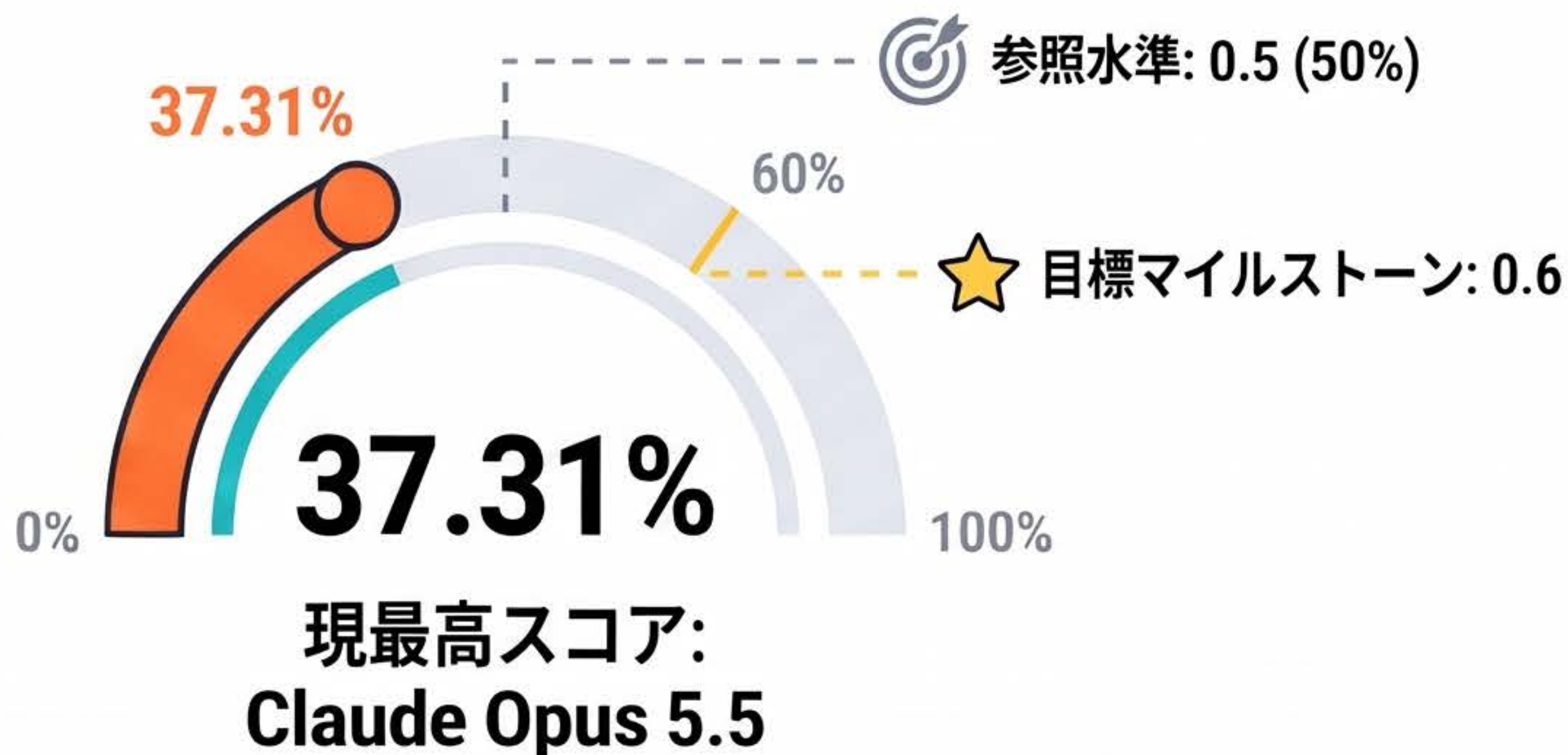
Vals RSI Indexは完全な「再帰的自己改善（RSI）」ではなく、「AIがAIを研究開発する能力」を測る長時間・オープンエンド型ベンチマーク。
現トップはClaude Opus 5.5（37.31%）だが、世代間ループの実証や完全な再現性には依然として重要な課題が存在する。

ベンチマークの構造と最新実績

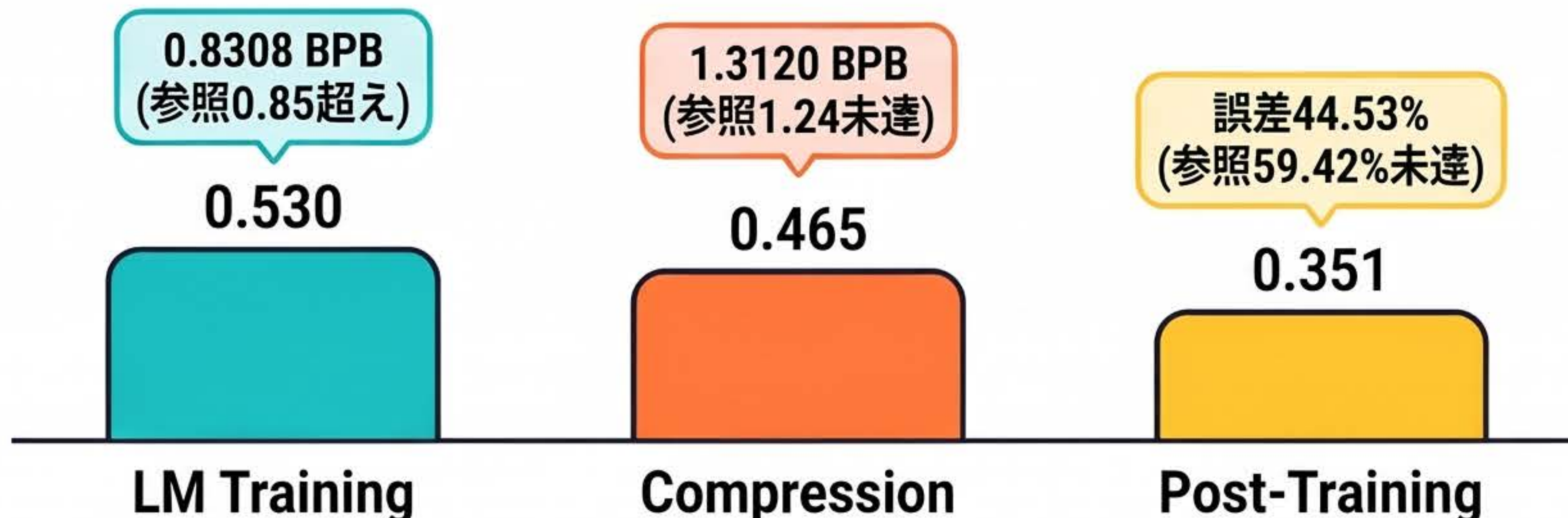


本質は「AI R&D能力」
の外部評価

再帰ループは未実装であり、
実質は自律的なAI研究開発能
力を測る指標。



Opus 5.5
課題別実績



誤解しやすい3つの技術的境界



再帰ループ（世代間改善）
は未検証

作成された後継モデルがさらに次の
モデルを作る検証は行われていない。



研究試行は1回のみで再現
性に制約

表示される誤差は評価時のばらつ
きで、研究の再実施分散ではない。



連続学習における自己劣
化のリスク

段階的な追加学習では破滅的忘却
により旧・新性能の双方が悪化。