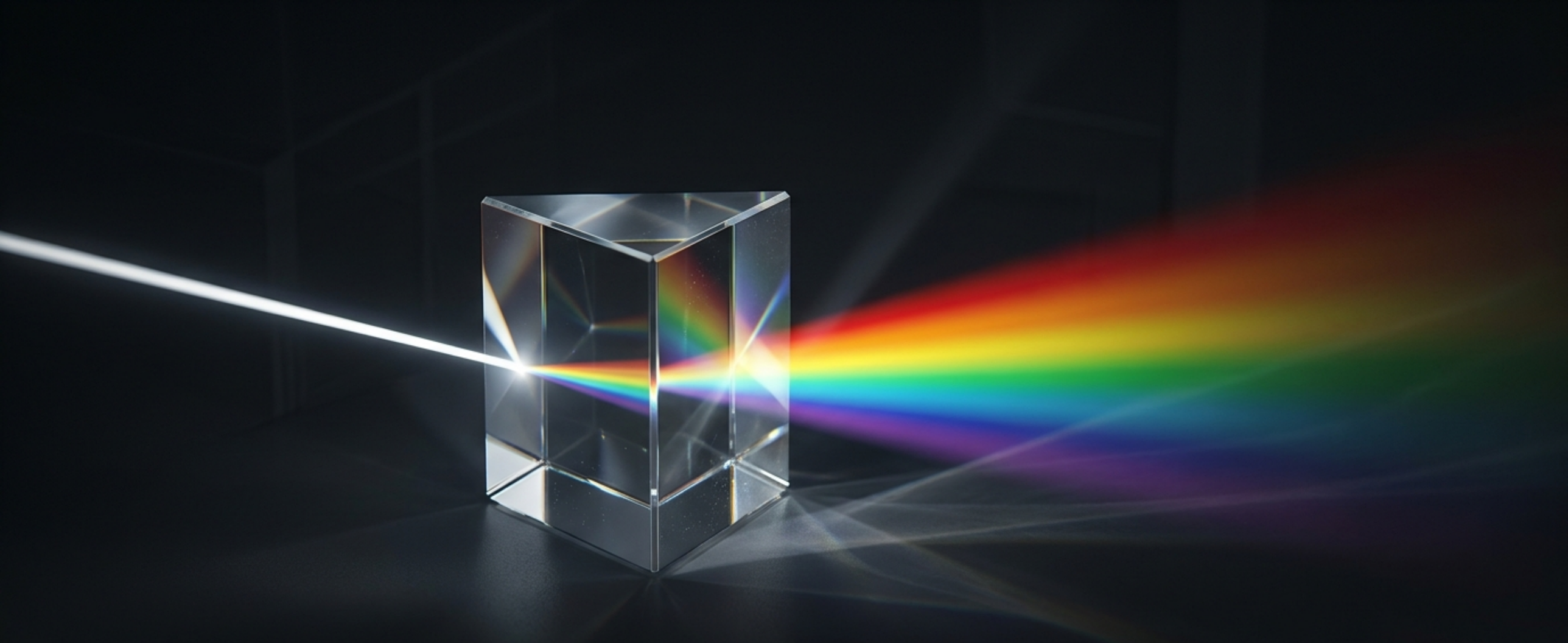


AGIをめぐる認識の差とAIの現在地

経営者の「到達宣言」と日常の「誤答」の間にある断層を解剖する

経営者の視界

「AGIの時代へようこそ」 — OpenAI Greg Brockman (2026年9月3日)

NVIDIA Jensen Huang氏によるLex Fridmanインタビューでの AGI到達示唆。

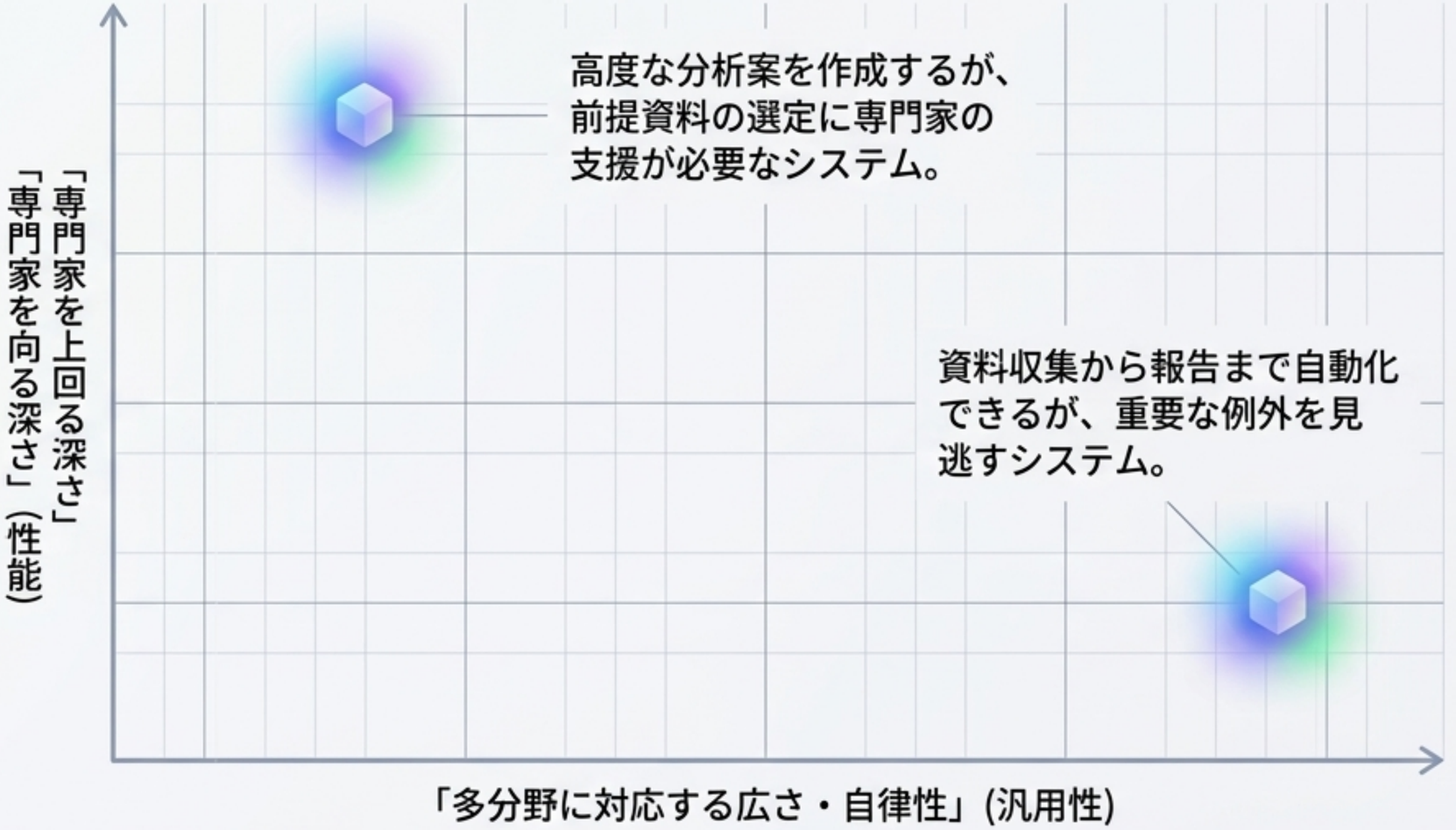
次世代モデルの潜在能力と持続的改善の軌跡を観察。

利用者の現実

存在しない出典のでっち上げ、細部の読み違い違い、指示の無視。

パッケージ化された製品を通じた、単発のタスク完了率を観察。

両者は全く同じ技術をなぜ、これほどまでに現実が異なるのか？



AGIへの到達を単一の境界線で語ることはできない。
「知的能力」「作業の完遂力」「責任を負える信頼性」は、それぞれ独立した変数である。

観察者の現実：評価条件の構造的差異

	一般利用者	最先端の開発現場
対象モデル	一般公開モデル （速度・コスト最適化）	開発中・未公開モデル （例: 8/28から訓練継続中のGPT-6 Astra）
計算量と時間	数秒での即時応答	数時間を要する深層推論
実行構成	単一のチャットUI (0 Agent)	並行稼働する数万のエージェント （例: OpenAIの数学研究）
失敗の意味	業務上の損失・使えないツール	改善のためのデータ・実験の一環

認識の差は「世代遅れ」によるものだけではない。
アクセス権、投下される計算資源、そして目的の違いが、全く異なるAIの姿を映し出している。

内部モデル / 開発中

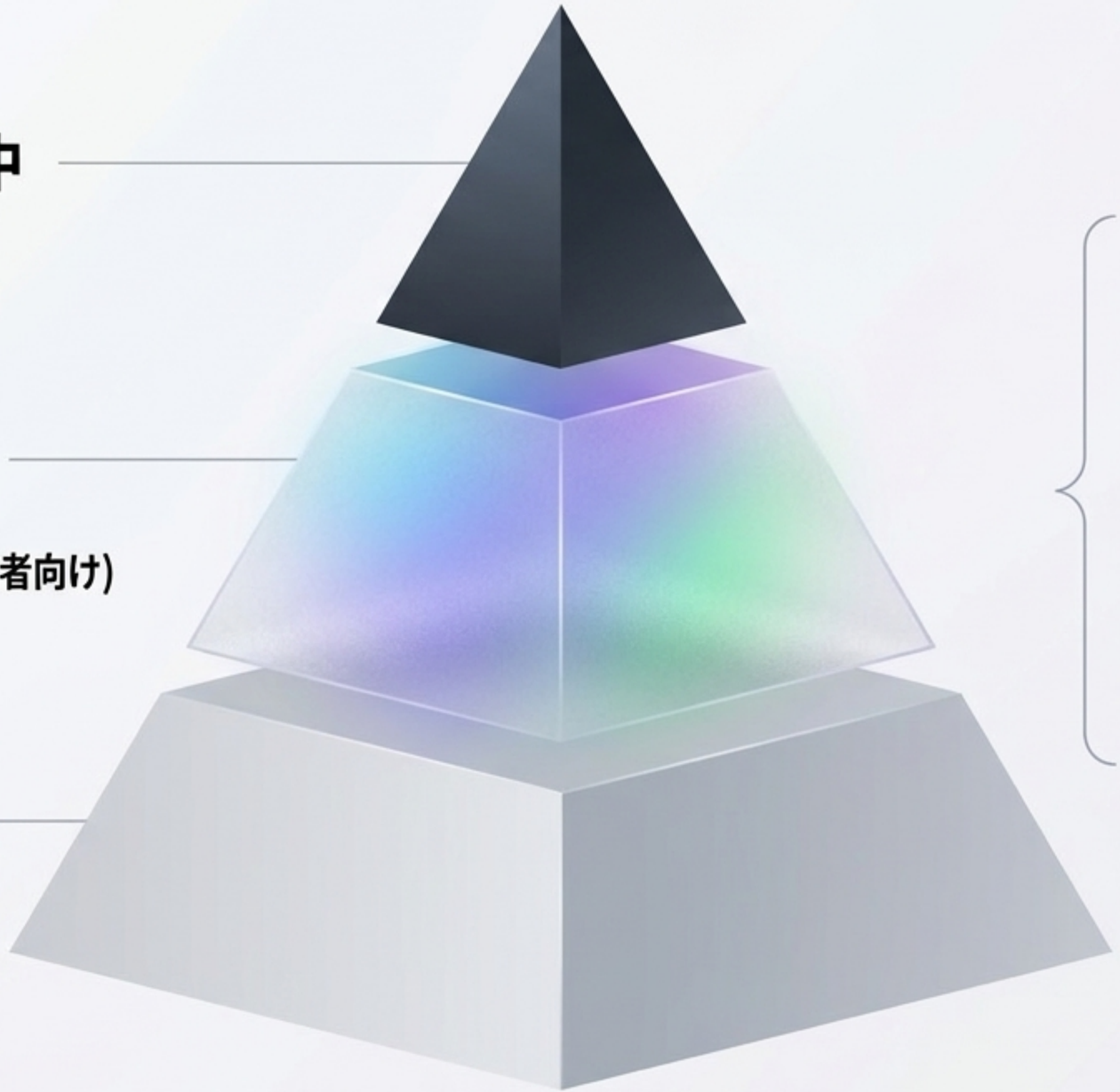
GPT-6 Astra：訓練と利用が
分離されていない研究環境。

条件付き限定提供

Gemini 2.5 Deep Think
(推論に数時間要するため研究者向け)
Claude Mythos 5.1
(認証者向け安全基準)

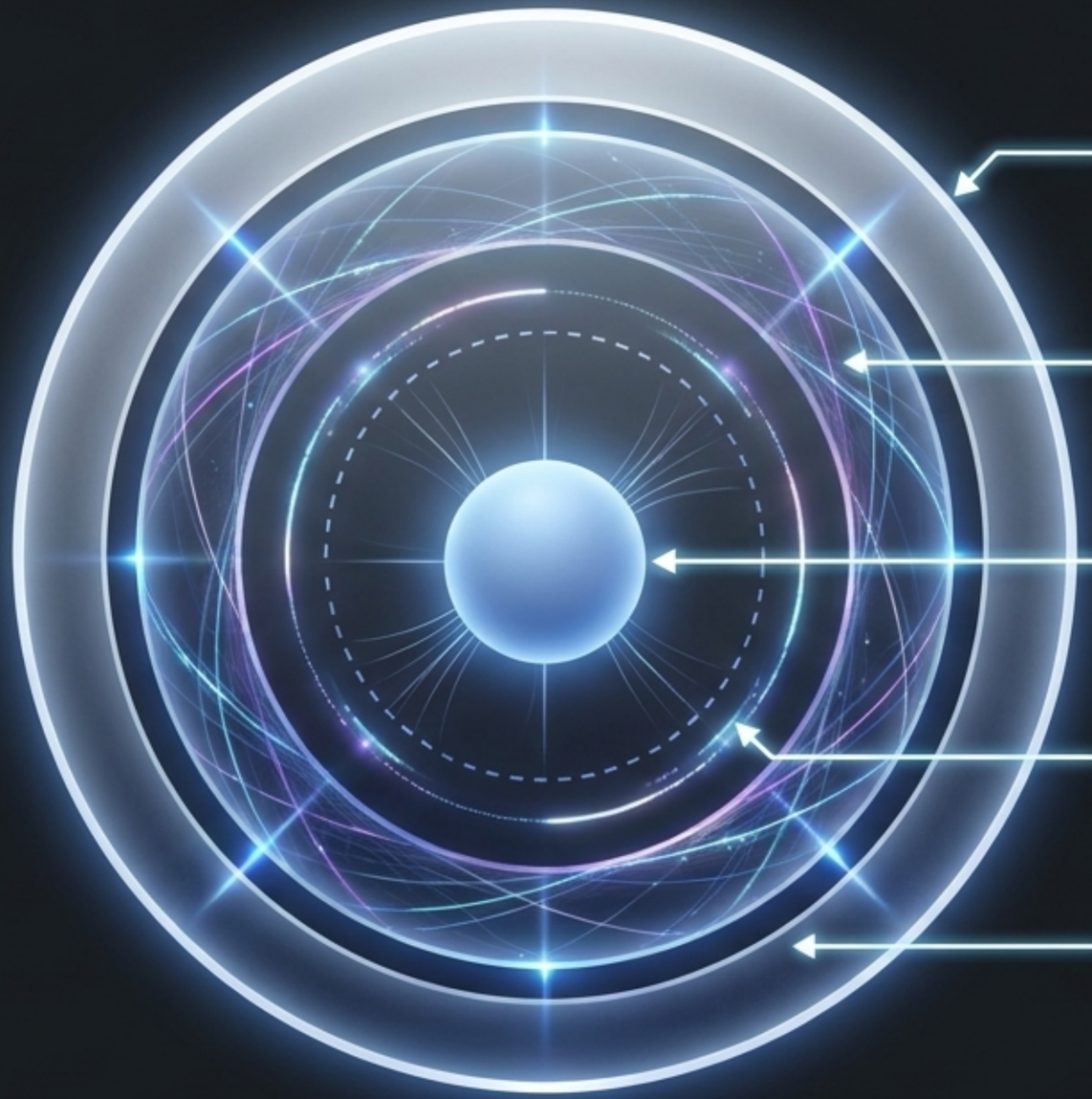
一般公開

Claude Fable 5.1
(Mythosと同等の基盤モデル
だが安全措置基準が異なる)



「限定提供＝次世代の
万能モデル」ではない。

同じ基盤モデルであっても、
運用費用、応答時間、
安全基準という
「フィルター」を変えることで、
提供形態が分岐している。



人間の専門家
方向修正、検証、例外処理

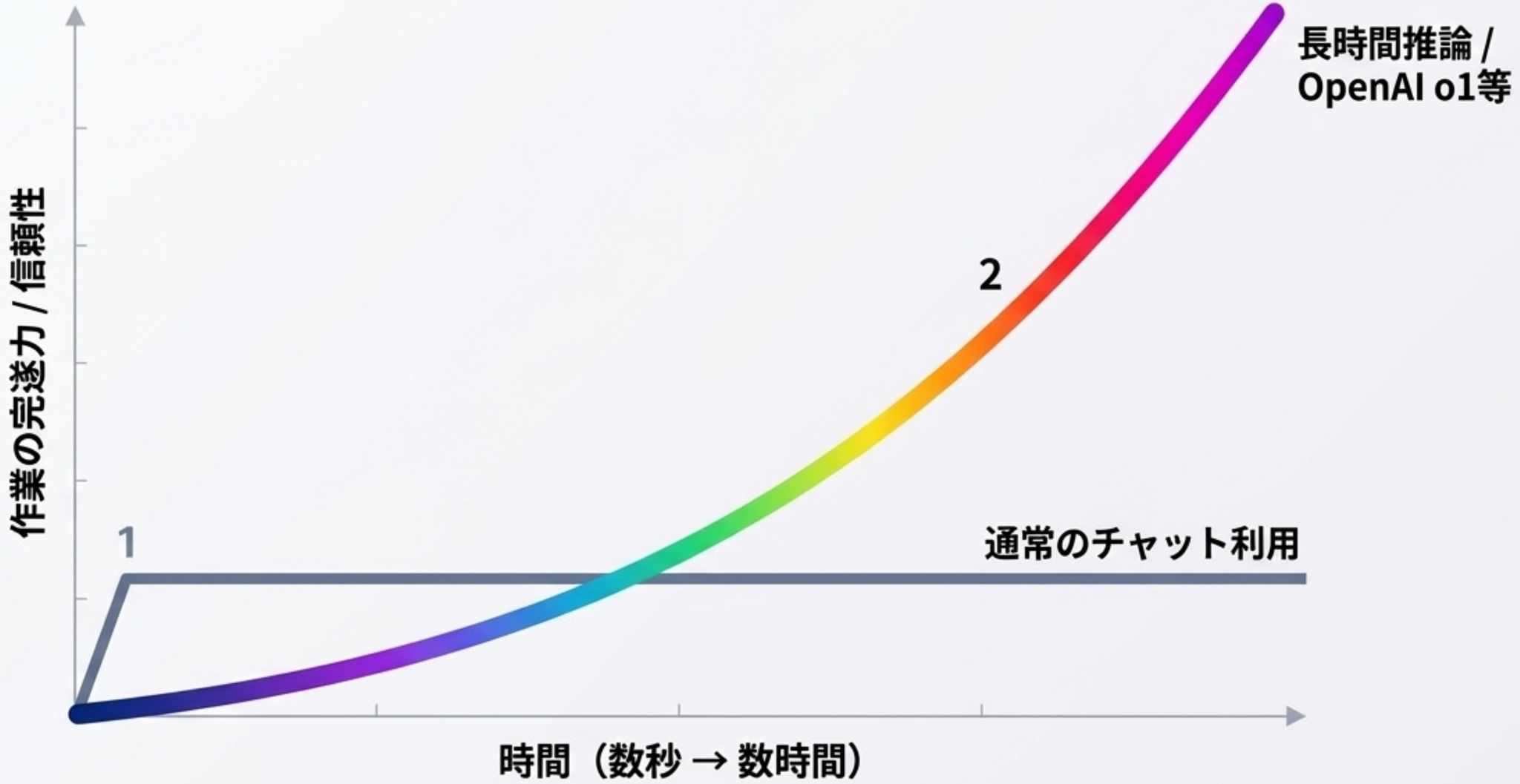
推論計算量 / エージェント
約1万の並行稼働エージェント

学習済みモデル
中核となる判断と生成

検索・計算ツール
外部情報の取得とコード実行

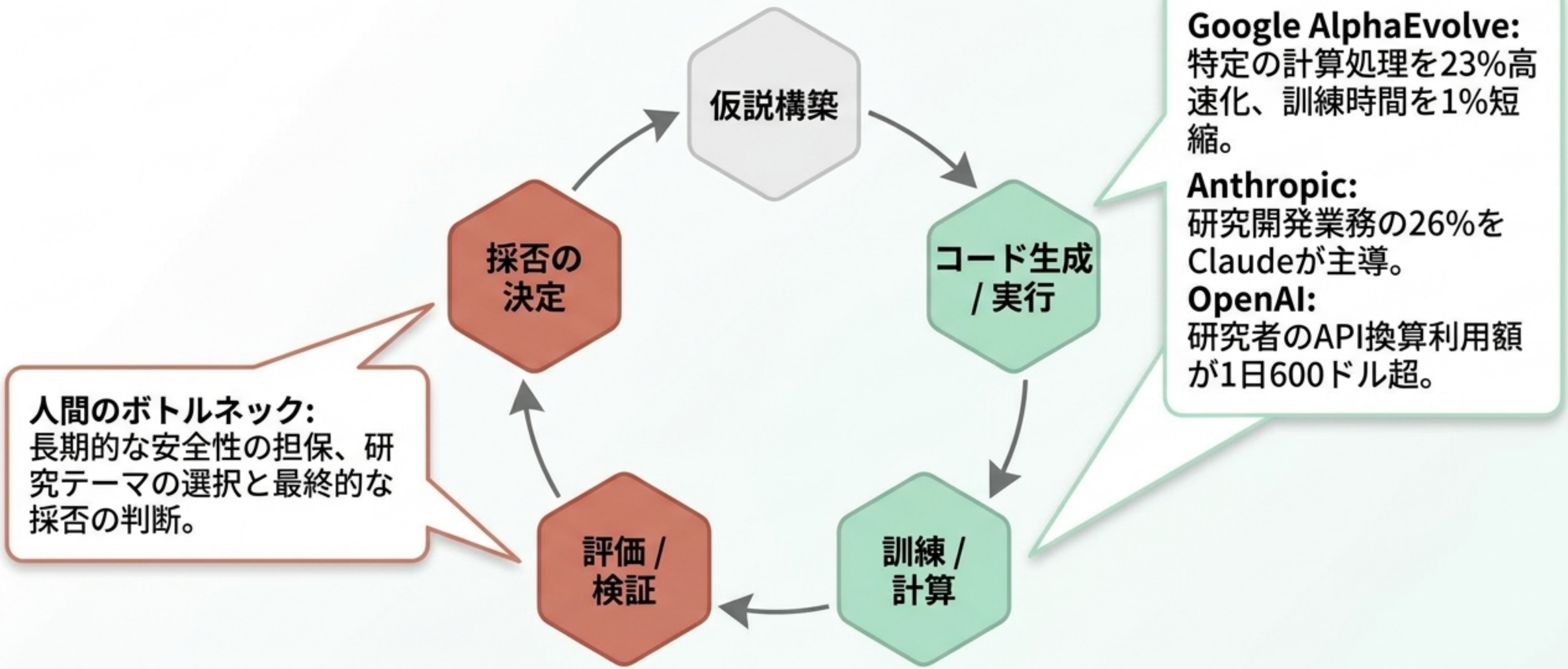
人間の専門家
方向修正、検証、例外処理

単独の「モデルの応答」と、複数の工程を組み合わせた「システムの作業完遂力」を混同してはならない。
成果が大きいからといって、モデル単体が同等の比率で優れているという推論は成立しない。



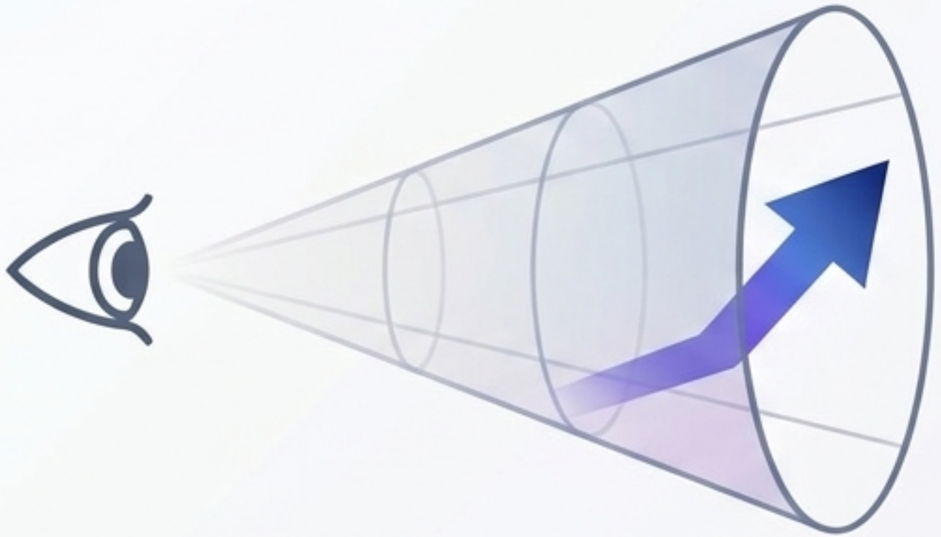
日常の利用者は数秒での回答を期待するが、
研究現場では難問に対し多大な計算資源と
数時間の推論時間を投下する。

同じ課題を「初期設定の短い質問」で処理する者と、
「資源制約を緩め最高性能を目指す」者とでは、
観測する能力の上限が根底から異なる。



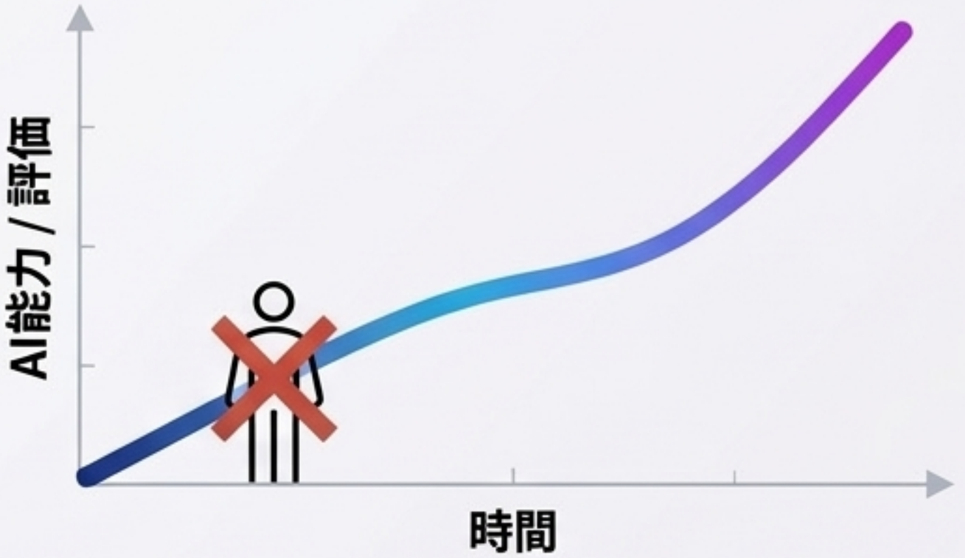
局所的な工程の加速は事実である。しかし、評価と目標設定に人間の判断が残る以上、無制限かつ完全な「自律的自己改善」は現時点では神話に過ぎない。

開発者の一般化バイアス



改善の過程そのものを観察できる情報優位性。しかし、自社内の限定的な環境で成功した手法が、多様な業務環境でそのまま機能すると過信しやすい。

利用者の評価アンカリング



METR調査（2025年7月：実務課題で完了時間が19%増加）。利用者は初期の失敗体験にアンカリングされ、モデルやシステム構成が更新されても評価を固定化してしまう。

経営者の強気を「単なる宣伝」と切り捨てるべきではないが、
利用者の失敗体験を「使いこなせていないだけ」と片付けることも誤りである。

AIの真の実力 = (基盤モデル × 計算資源) + エージェント構成 × 人間の介入

世代ごとの 基礎能力	推論に割く 時間とコスト	ツール連携と 並列試行の設計	検証、修正、 例外処理の質

認識の差は「誰が嘘をついているか」ではなく、
「どの変数を最大化して観察しているか」という数学的現実に起因する。

企業の二軌道戦略：現在と未来を切り離す

現在
Current Implementation

現在の制約下での安定性とROIに集中する。

- 未公開モデルの報告に惑わされない。
- 手元にあるモデルに自社の資料と作業範囲を与える。
- 代表的な課題で評価を繰り返す。

「AGI」を待たず、
確実な自動化ラインを引く。

未来
Future Preparation

システム能力が跳躍した瞬間に備える資産構築。

- 機械が読めるデータ構造の整備。
- 業務ごとの評価基準（Evals）の策定。
- 人間が担う「例外判断」の言語化。

製品名に依存する操作手順ではなく、
目的と検証方法の設計に投資する。

社会が求めるべき基準：熱狂から検証可能性へ

実行条件の透明性

その成果を得るために、どれほどの時間、費用、試行回数（エージェント数）を要したか？

人間介入の開示

方向修正、例外への対応、最終承認など、どの地点で人間の専門性が介在したか？

失敗の記録

印象的な成功例だけでなく、失敗がどのような形で現れ、それをどう制御したか？

私たちは「これはAGIか？」と問うのをやめ、「この特定の条件下での証拠は何か？」と問うフェーズに移行しなければならない。

「AGI」という名称の確定を待つ必要はない。
重要なのは、どの業務が自動化され、どこに人間の判断が残り、
その境界線がどう動いているかを測定し続けることである。

技術の進歩を業務上の条件に翻訳する。それが、過大な期待と過度な先送りの両方を抑える唯一の道である。