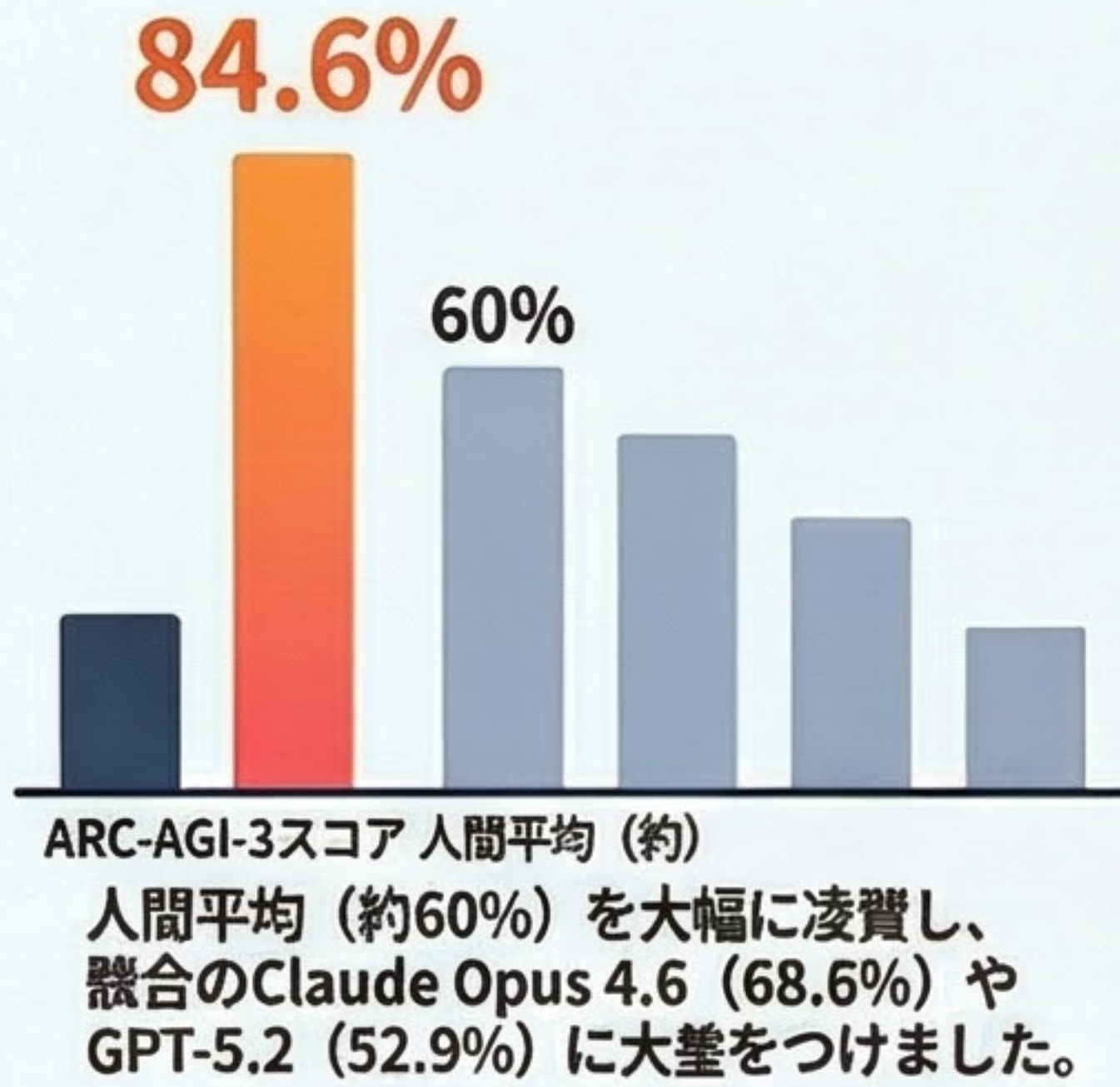
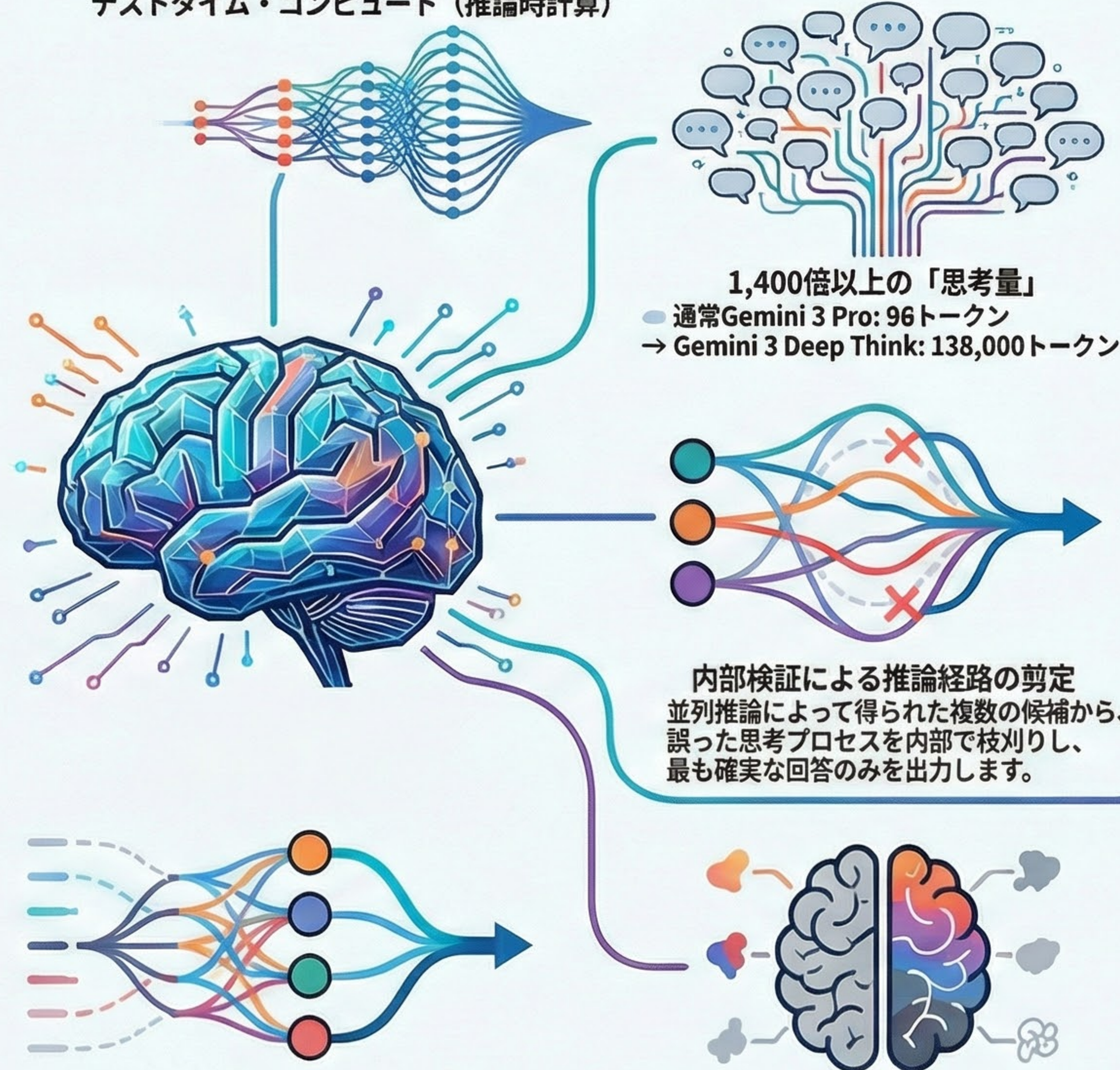


Gemini 3 Deep Think：推論AIの到達点と「知能の代償」

圧倒的な推論・抽象知能の証明



「考える時間」：知能を支えるメカニズム テストタイム・コンピュート (推論時計算)



内部検証による推論経路の剪定
並列推論によって得られた複数の候補から、
誤った思考プロセスを内部で枝刈りし、
最も確実な回答のみを出力します。

依然として残る「非均一な能力」
機械推論には強い一方、ソフトウェア工学
やマルチモーダル処理、創造的文書執筆で
は統合モデルに劣る場合があります。

「知能の代償」と効率性の壁



Model	ARC-AGI-2 スコア	推論コスト (1タスク)
Gemini 3 Deep Think	84.6%	\$13.62 (Very High)
Claude Opus 4.6	68.8%	(データなし)
GPT-5.2	52.9%	~\$1.90 (Medium)
Gemini 3 Pro	31.1%	\$0.81 (Low)
人間平均	~60.0%	\$0.42 (Goal)

精度とコストのトレードオフ：GPT-5.2は精度(52.9%)で定まるもののコストは約1.90ドルであり、実用的なコスト効率の面ではDeep Thinkは線路を逸しています。

AGIへの道程と次なる課題

2026年3月「ARC-AGI-3」リリース予定
現行ベンチマークの飽和を脱し、より実世界に近い「エージェンシー」を試す諸機構へと移行。



オーバーフィッティングの
既知：推論プロセスに特定の
カラーマッピングが検出されて
おり、高いスコアの一部が学習
データの記録に由来する可能性。



人間の監視が必要な領域：
推論が過剰でもハルシネーシ
ョンは完全には排除できず、金融
判断や研究分野では継続として
人間のチェックが不可欠です。