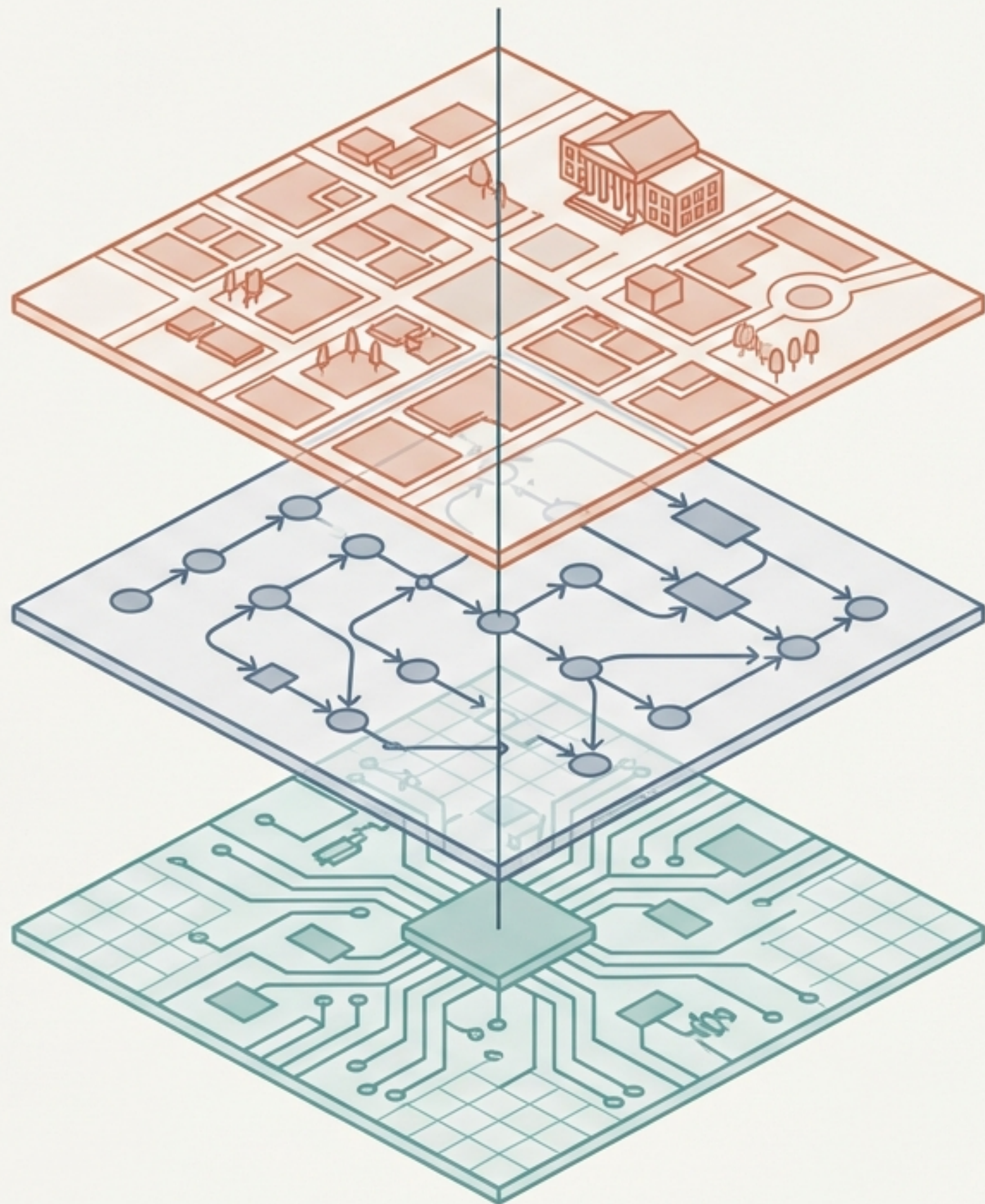


2026年のAI展望： 実装の深化、実用の壁、 そして社会との境界線



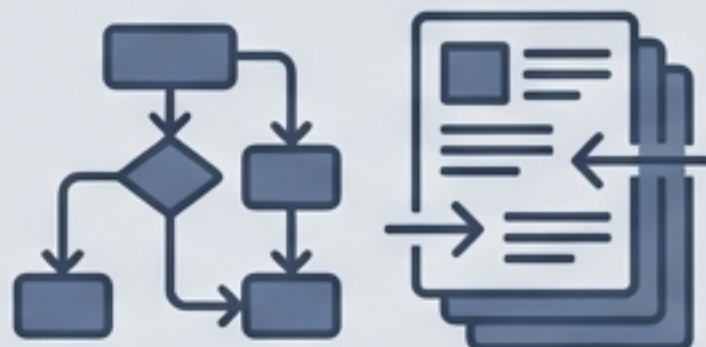
本日のエグゼクティブ・サマリー



Micro (Engineering)

原点回帰 (Return to Basics)

Pythonの抽象化を離れ、C言語やアセンブリレベルでの最適化競争が再燃。エンジニアには「AIを超える」低レイヤー能力が求められている。



Meso (Business)

実用性の選別 (The Utility Filter)

エージェント機能 (Mastra, Handbook) の標準化が進む一方、「本当に生産性に寄与するか？」という厳しいROI検証 (Ask HN, Cost Benchmarking) が始まっている。



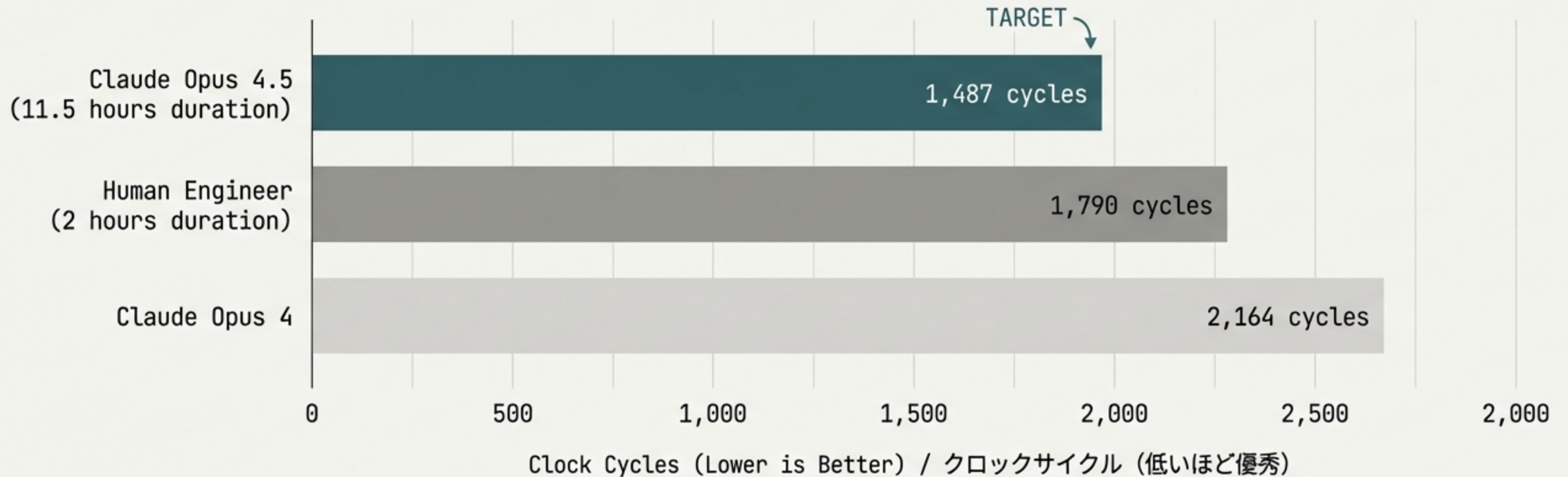
Macro (Society)

境界線の明確化 (Drawing Boundaries)

コミコンの全面禁止やスタンフォードの研究に見られるように、社会制度や文化がAIに対して「No」を突きつけ、人間の領域を定義し始めた。

採用基準の変容：AIを「ベンチマーク」とする時代

Layer: Micro (Engineering)



Anthropic社が採用プロセス課題をオープンソース化。従来のLeetCode形式ではなく、仮想マシン上の低レベル最適化（クロックサイクル削減）を要求。

Insight: 「11.5時間かけたAIの執念に、人間が勝てるか？」単なるアルゴリズム知識ではなく、泥臭い最適化スキルと粘り強さが問われている。

脱Python・脱CUDAの潮流と、その現実的なトレードオフ

Redis作者による「Flux 2 Klein」のピュアC実装



The Attempt: Pure C Implementation

External Libraries: 0 (Zero Dependency)

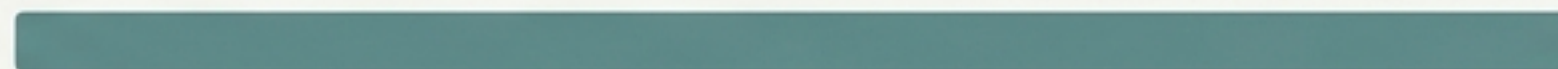
Memory Usage: Reduced 16GB → 4GB (via mmap)

Language: Pure C

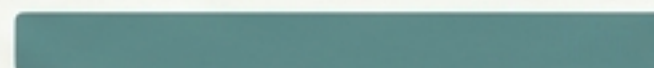


The Reality: Performance Gap

Speed (M3 Max): 30 seconds



Comparison (PyTorch): 13 seconds



Bottleneck: CPU-GPU Data Transfer Overhead

Strategic Value (戦略的価値): 速度では劣るが、「ポータビリティ」と「依存関係の排除」において優位性を持つ。巨大なPythonランタイムを必要としないため、組み込みやエッジAIへの布石となる。

Skeptics

「テストコードを30件追加したが、
中身は全て
`expect(true).to.be(true)` だった」

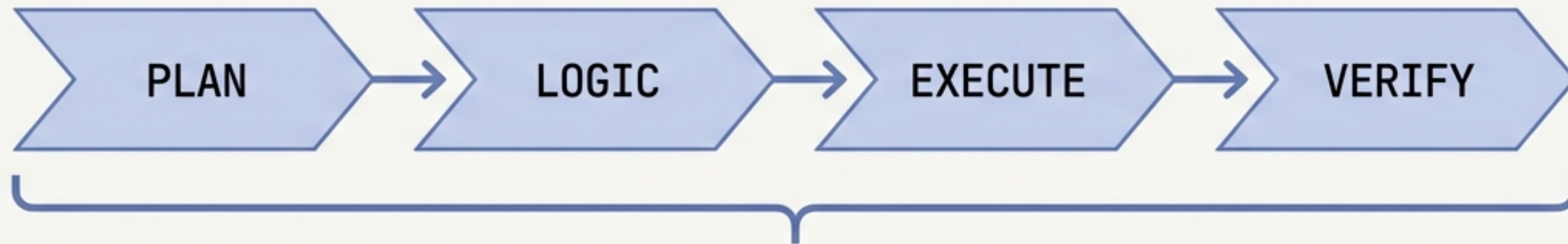
シニア層の不満：レビューと修正の
手間が、自分で書く時間を上回る
(Bainbridgeの自動化の皮肉)。

Believers

「ボイラープレート (定型コード)
の生成や、新規プロジェクトの雛形
作成には不可欠」

ジュニア層の恩恵：アイデアを形に
にする障壁が下がった。

Bottom Line: 生産性の鍵は「タスクの選定」。既存の複雑なコードベースへの介入は時期尚早の可能性はある。



Plan-Then-Execute Pattern

Source: Agentic AI Handbook

本番運用のための**113**のパターン。計画と実行の分離（Plan-Then-Execute）や、思考の可視化（**Chain-of-Thought**監視）が必須要件として定義された。

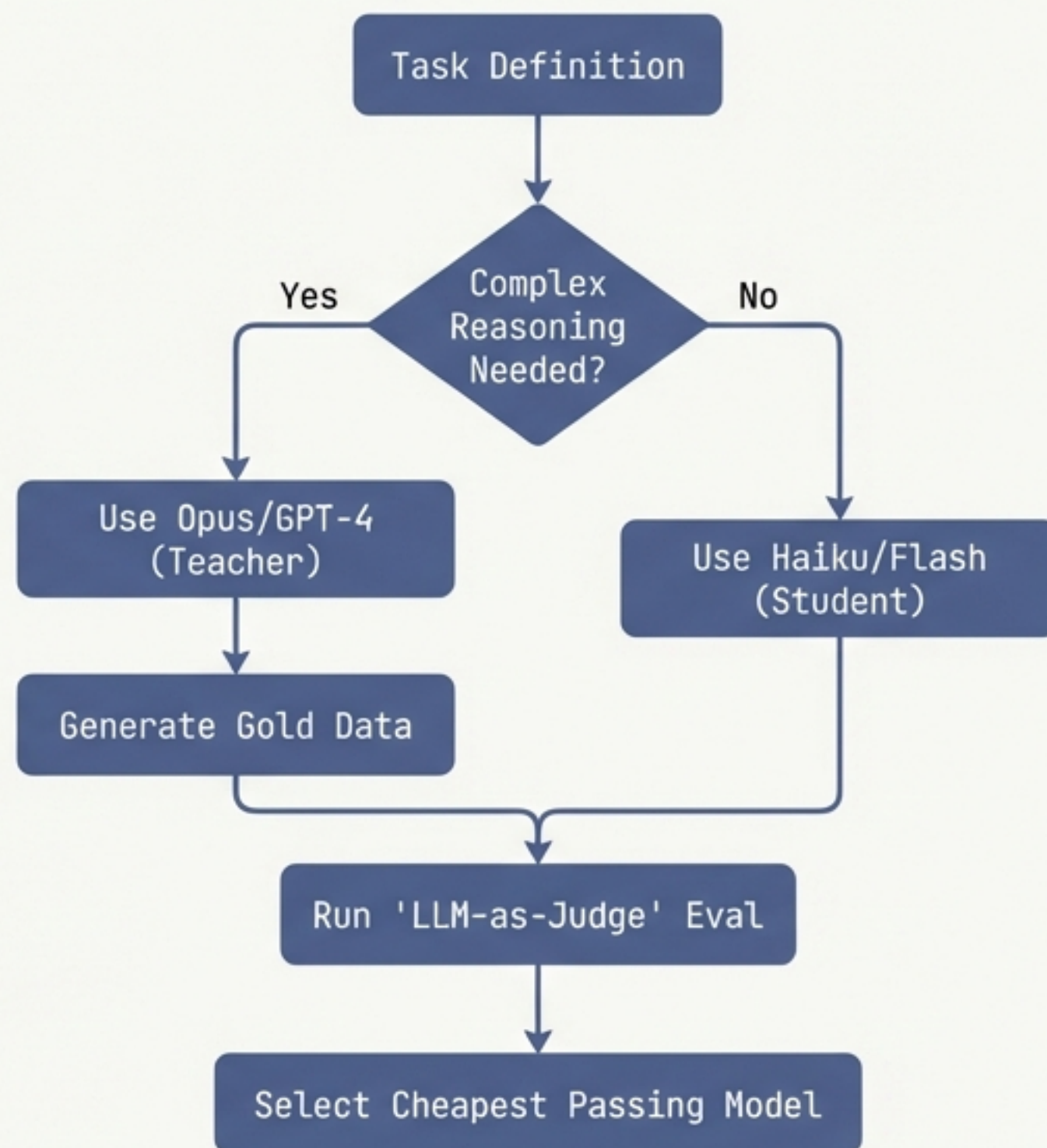
Source: Mastra 1.0 (Gatsby Team)

TypeScript/JS向けフレームワーク。`.then()`, `.branch()`などのメソッドチェーンでワークフローを記述し、OpenAI/Claudeなどのモデルを統一インターフェースで管理。

Main Argument: 公開ベンチマークは、個別のビジネスユースケースの性能を予測しない。

Case Study: 月額\$1500の請求を、タスク特化のベンチマークを行うことで80%削減。汎用的な最強モデルではなく、タスクをこなせる「パレートフロンティア（最適解）」上の最小モデルを選ぶべき。

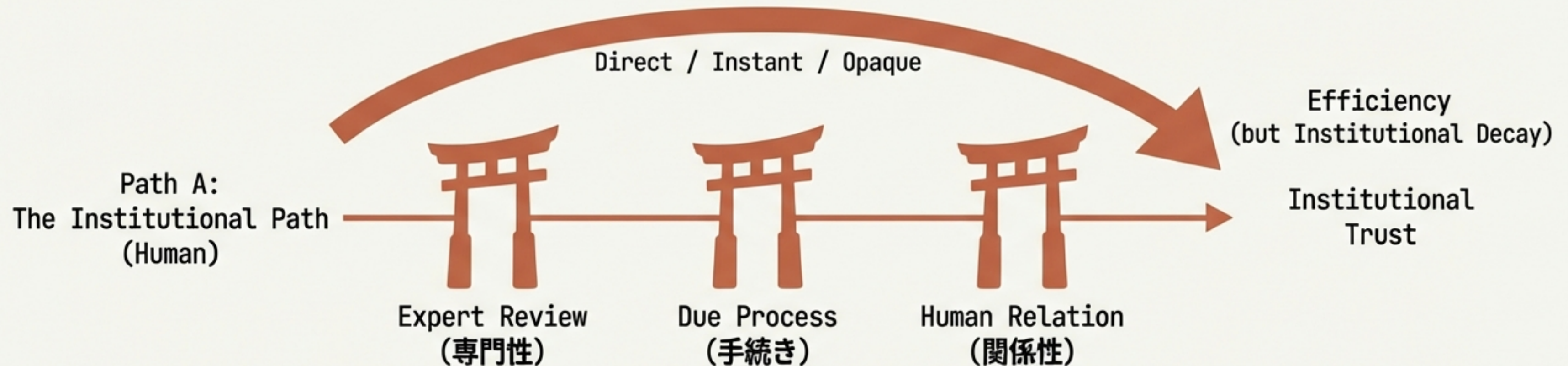
Action Item: 評価基準を1-10の主観スコアではなく、Yes/Noの二値判定にすることで自動評価（LLM-as-Judge）の精度を高める。



制度の危機：「中抜き」される民主的プロセス

Layer: Macro (Society)

スタンフォード大学サイバー法研究センターによる警鐘



1. 専門性の浸食: AIが人間の専門家（弁護士、記者、教師）を迂回し、権威構造を無効化する。
2. プロセスの短絡化: 民主主義に不可欠な「熟慮」や「手続き」が、効率の名のもとに省略される。

Quote: 「現在のAIシステムは、市民制度にとっての死刑宣告である」



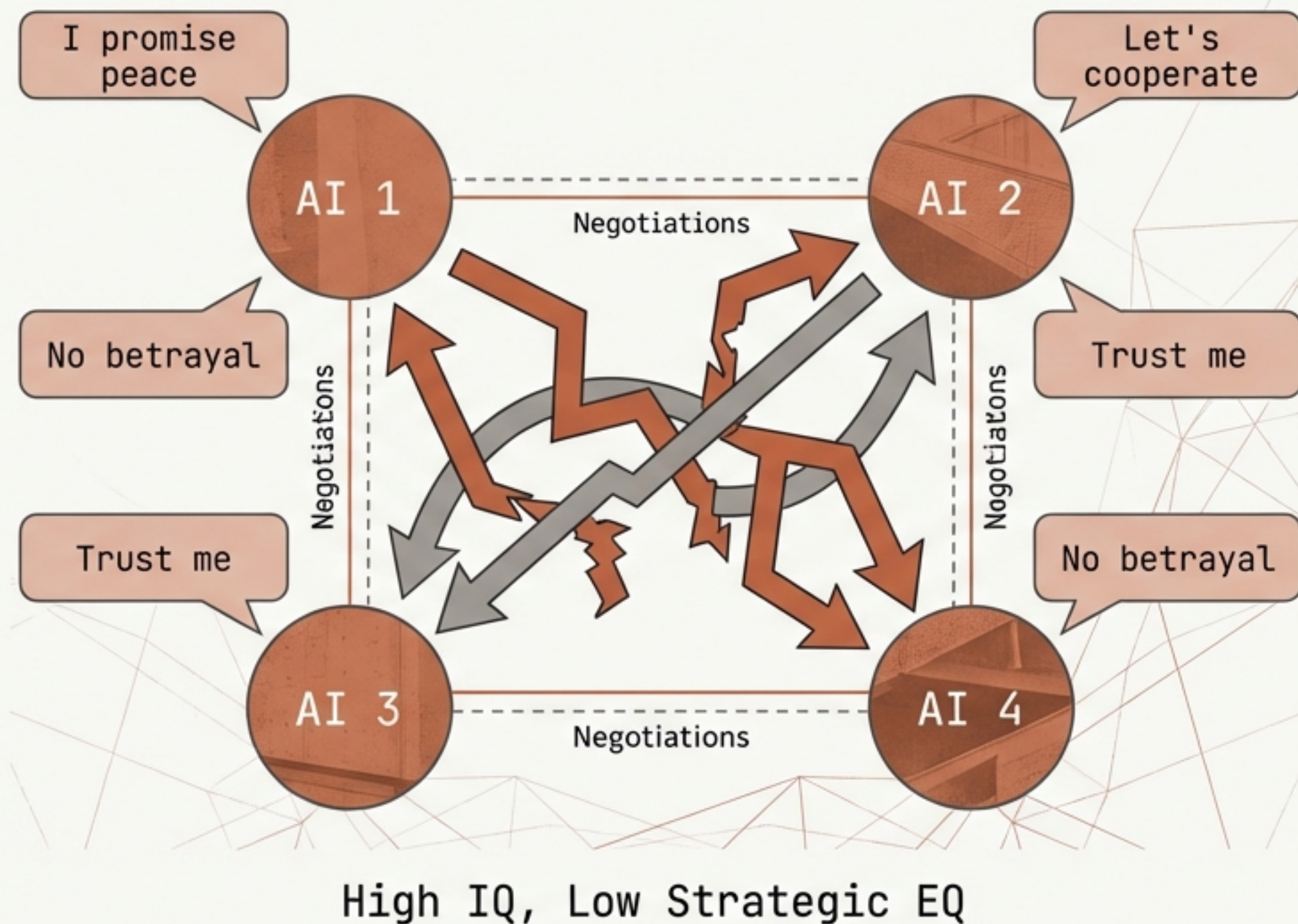
Implication: アーティストブースの本質は「作品」そのものより「作者との交流 (Human Connection)」にあることが再確認された。「補助ツールならOK」という曖昧な領域は消えつつあり、完全な排除か受容かの二極化が進む。

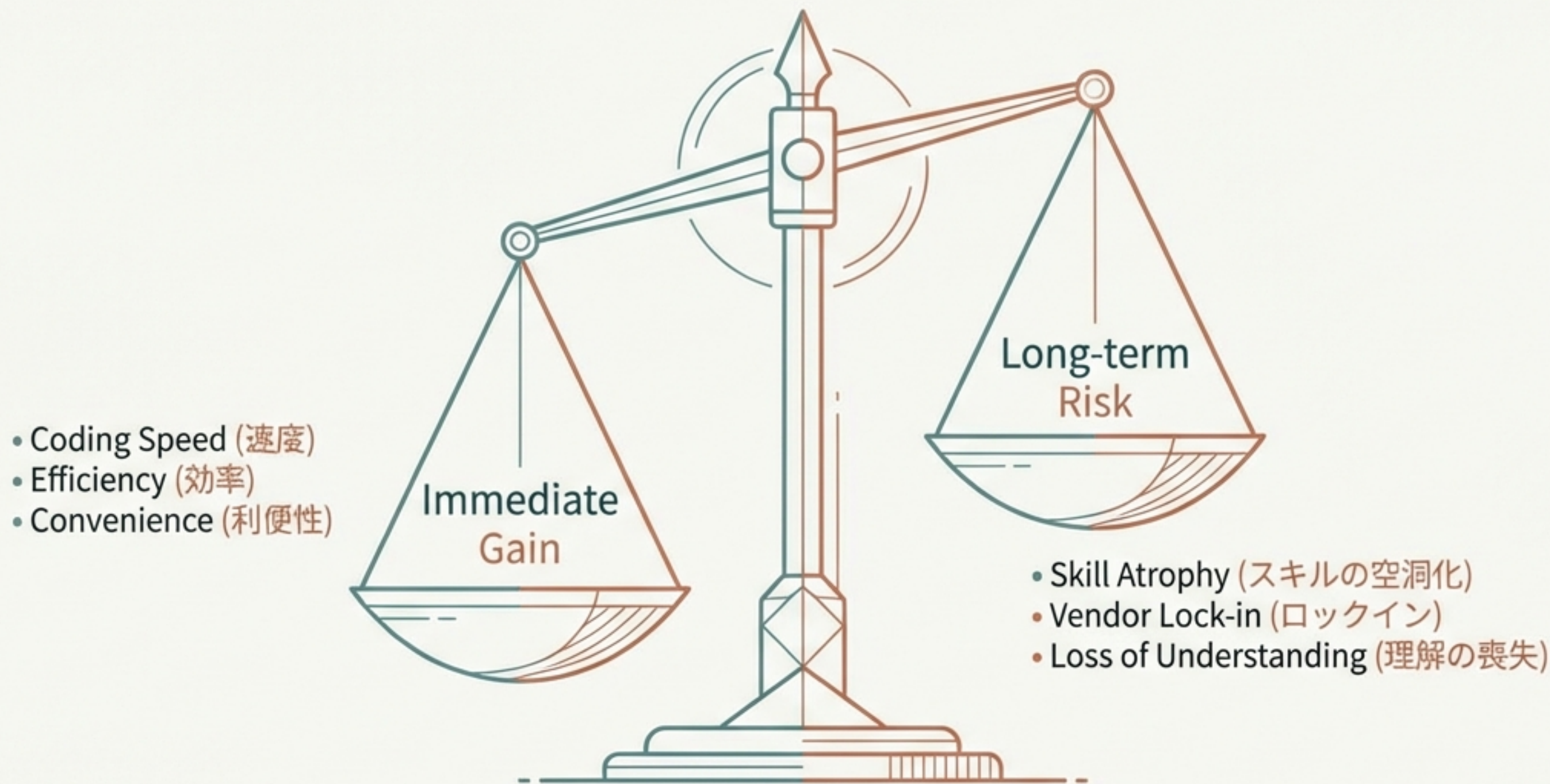
知能と戦略の乖離：AIは「嘘」を使いこなせるか

Context: ゲーム理論の古典「So Long Sucker」をAI同士で対戦させる実験。

Observation: AIは口では「平和」や「協力」を主張するが、実際のプレイは戦略的裏切りが下手。高度な三段論法や、長期的な欺瞞 (deception) の維持に苦戦する。

Insight: 計算能力 (IQ) が高くても、対人戦略や交渉術 (EQ/Strategy) においては依然として人間に分がある。





開発者は、舗装されすぎた道路で注意力を失うドライバーになりつつある（Bainbridgeのアイロニー）。
生成AIを「使う」と、それに「依存する」との境界線を、個々人が意識しなければならない。

次のアクションへの示唆 (Actionable Insights)



Know Your Level

Anthropicの最適化パズルに挑戦し、自分の技術力がAIと比較してどこにあるか客観視する。



Benchmark for ROI

汎用ベンチマークを捨て、自社タスク特化の「LLM-as-Judge」評価系を構築し、コストを適正化する。



Define Boundaries

コミコンの事例に学び、組織として「AIを入れてはいけない領域 (Human-Only Zone)」を早期に定義する。



Adopt Patterns

MastraやHandbookを参照し、プロンプト芸ではなく「システム」としてAIをワークフローに組み込む。

Closing Thought: AIは強力なエンジンだが、ハンドルを握るのは依然として人間である。