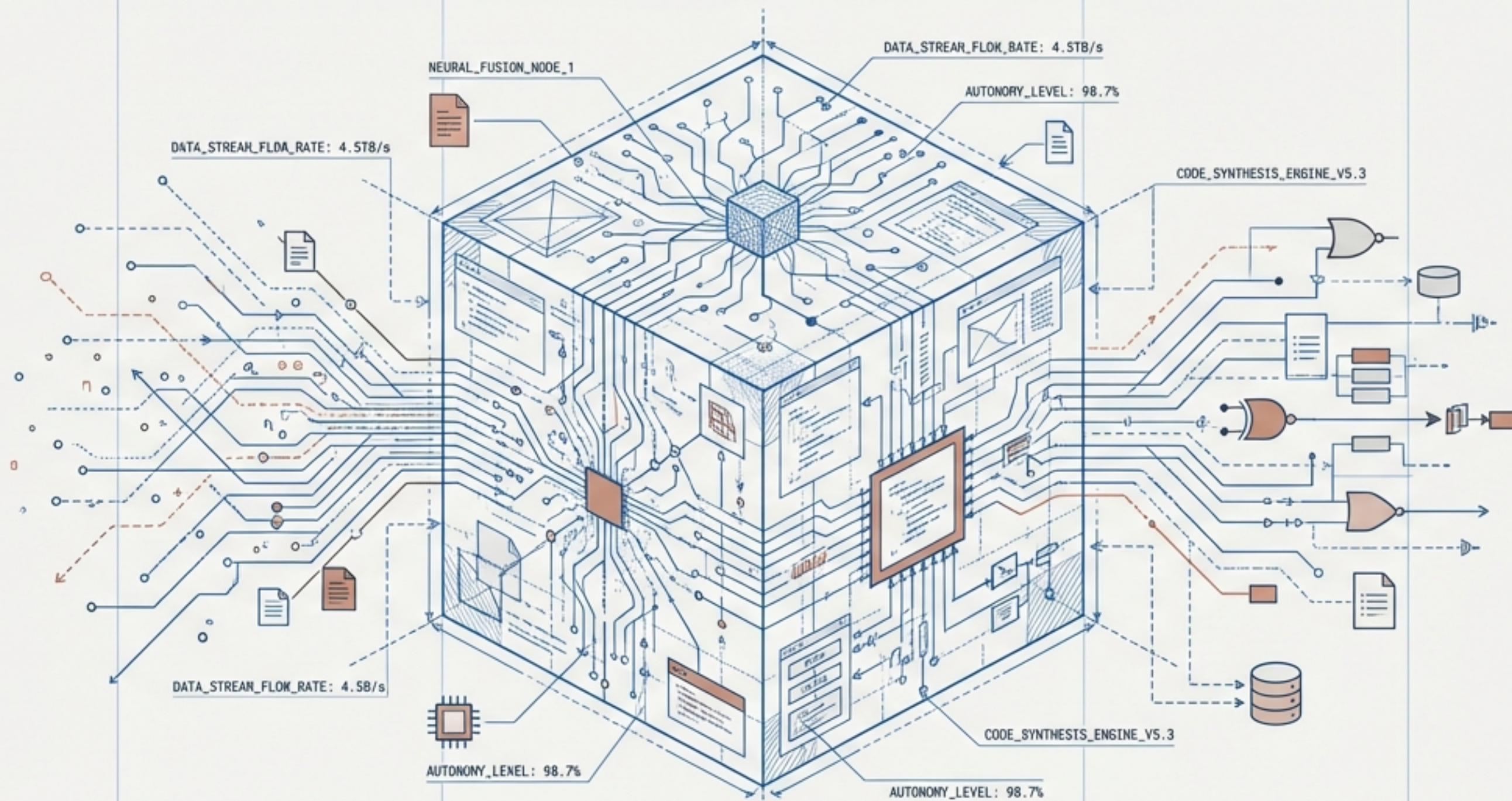


# GPT-5.3-Codex 総合調査レポート

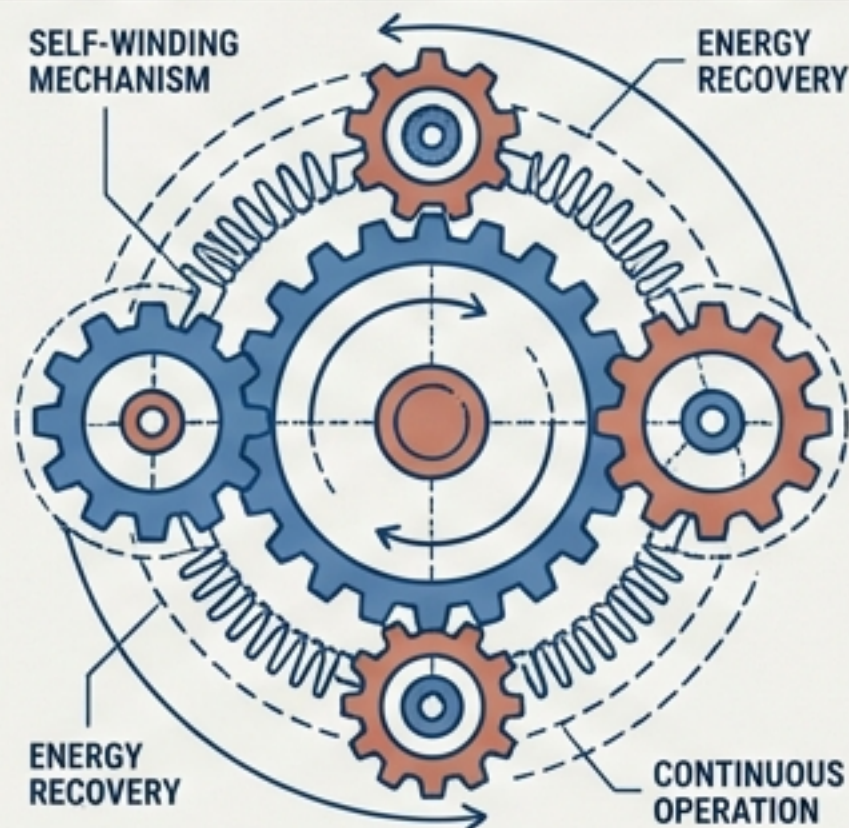
「コード生成」から「自律型開発エージェント」への進化的飛躍と、Claude Opus 4.6との比較





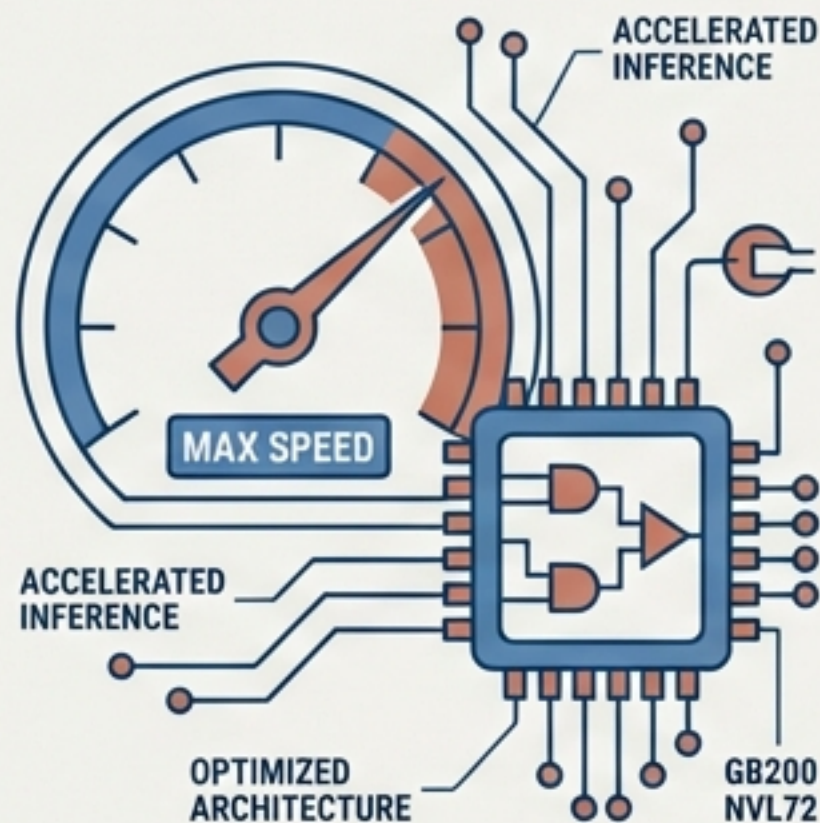
# エグゼクティブ・サマリー：AIは「ツール」から「同僚」へ

GPT-5.3-Codexは、従来の「コード補完」を超え、長時間自律稼働する「エンジニアリング・エージェント」へと進化した。



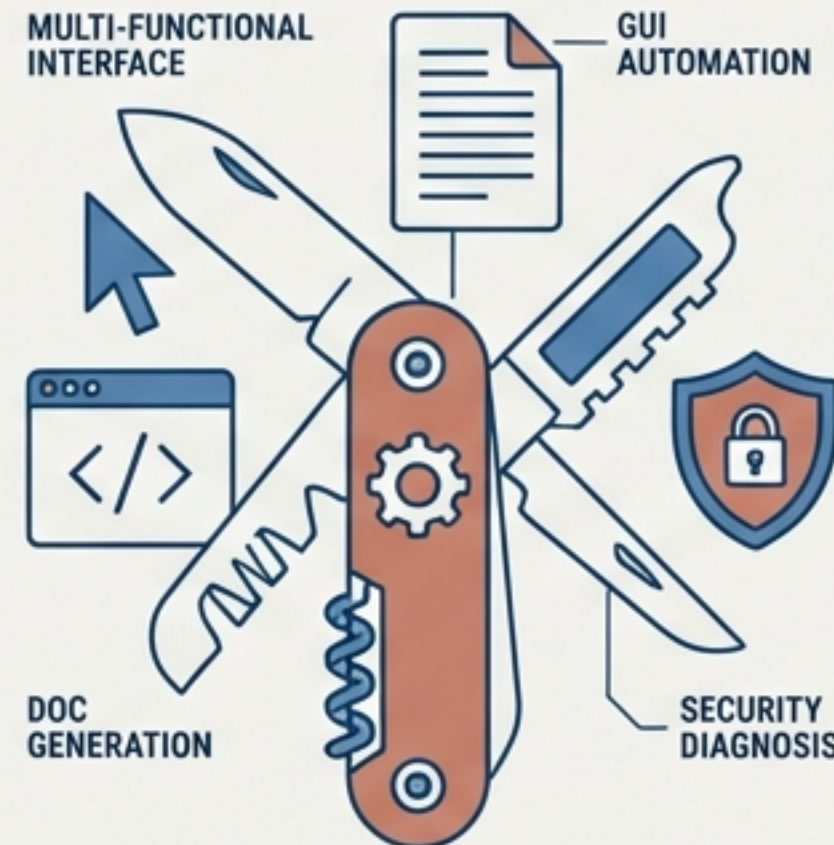
## 圧倒的な自律性

数百万トークンに及ぶコンテキストを維持し、数日間のタスクを遂行。  
「自ら計画・実行・修正」が可能に。



## パフォーマンス

推論速度25%向上、Terminal-Bench 77.3%（業界最高）。NVIDIA GB200 NVL72システムによる最適化。



## 汎用性

コーディングだけでなく、GUI操作（OSWorld 64.7%）、ドキュメント作成、セキュリティ診断までカバー。



# パラダイムシフト：「Copilot（副操縦士）」から「Agent（自律的な職人）」へ



## GPT-5.2-Codex (Past)



- 人間が指示し、AIが書く
- 単発の関数生成
- 短期間の記憶
- 成果物: コードの断片 (Snippets)



## GPT-5.3-Codex (Future)



- AIが計画し、人間が承認する
- プロジェクト全体の把握
- 長期記憶と自己修正
- 成果物: 動作するソフトウェア (Working Software)

特筆事項：モデル自身が自身の学習プロセスのデバッグを行った（OpenAI内部事例）



# 技術的進化：NVIDIA次世代GPUと統合された知性

Context Efficiency: UP

## Model Architecture

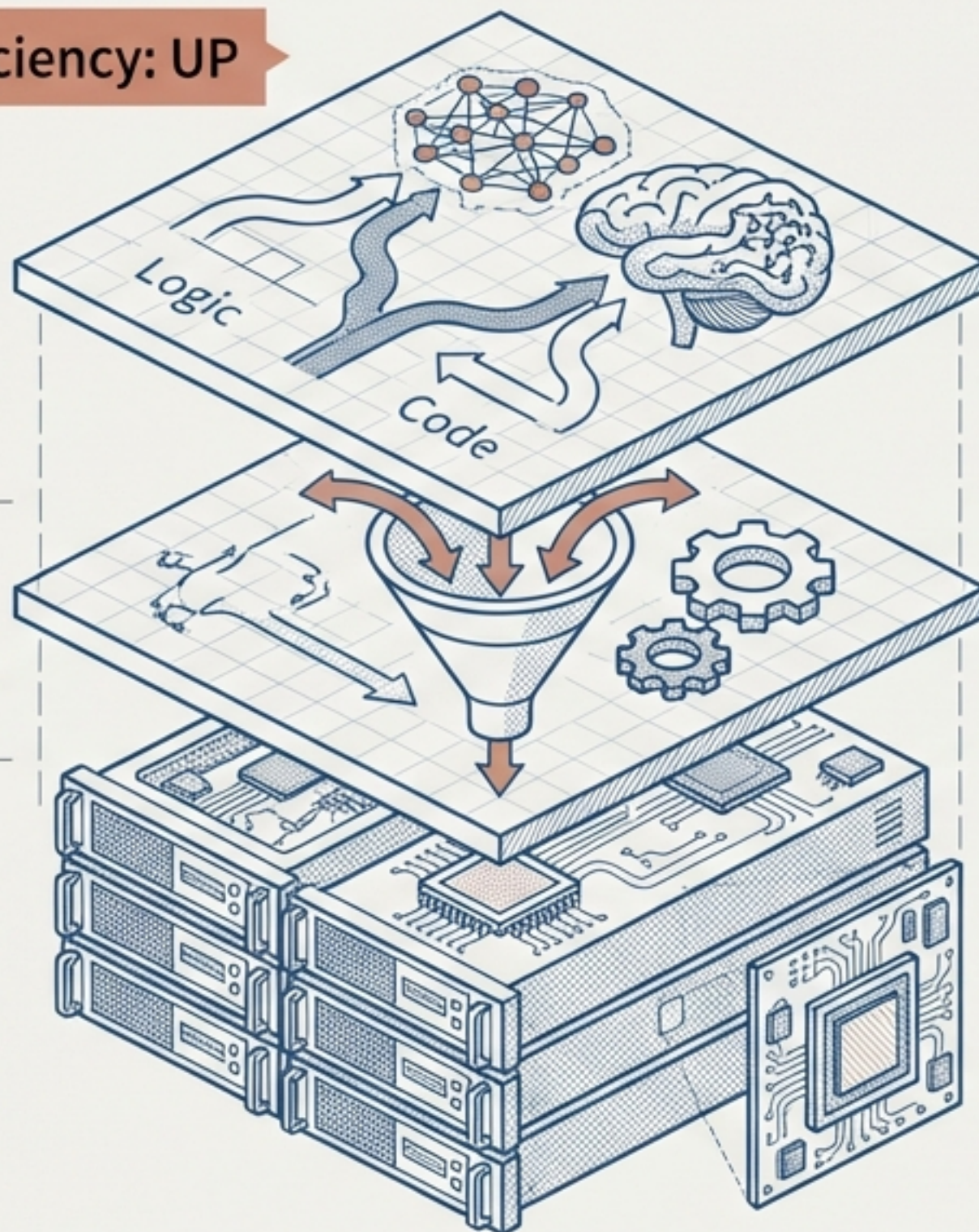
汎用モデルの推論力  
+ Codexの実装力

## Optimization

推論速度 25% 向上 / コスト効率化

## Hardware

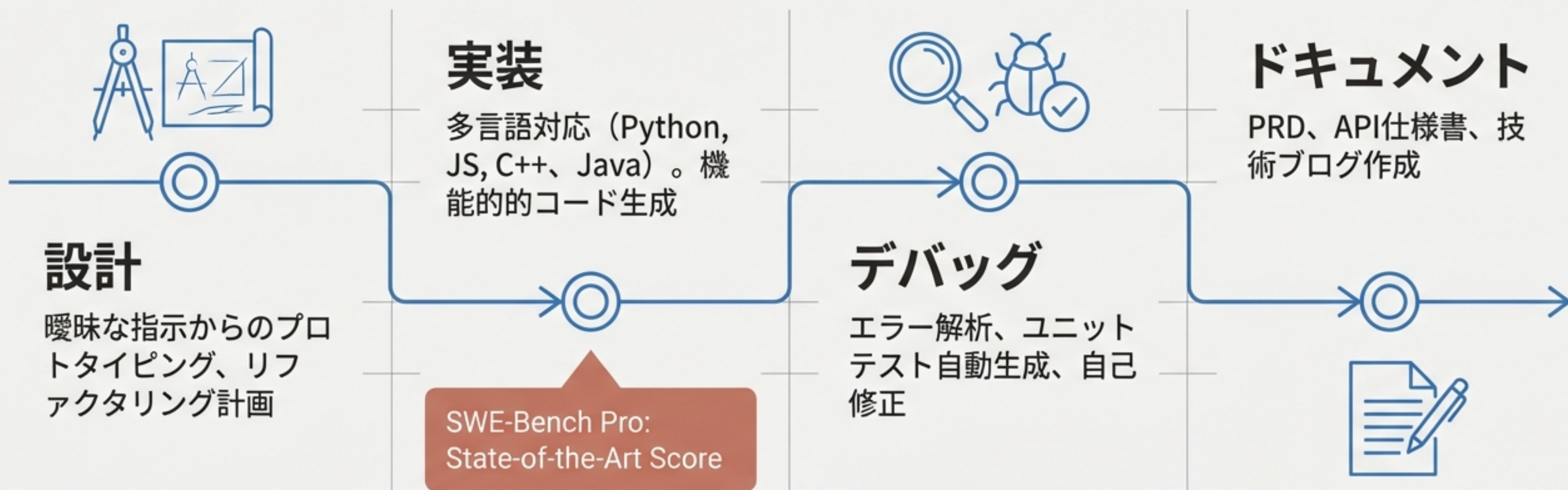
NVIDIA GB200 NVL72 System



Latency: -25%



# 守備範囲の拡大：要件定義からデプロイまで





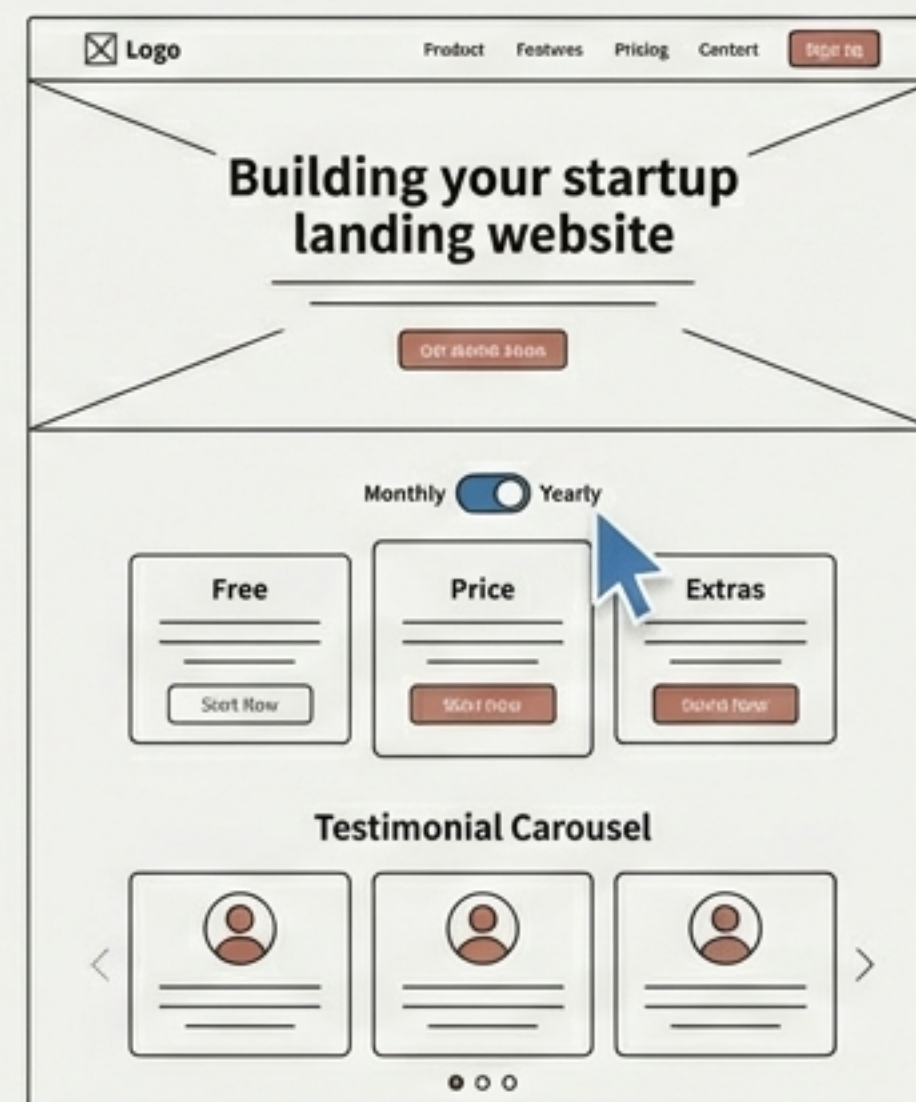
# 拡張能力：スクリーンを「見て」、マウスを「操作」する

OSWorld Score: 64.7% (Human: 72%)

## INPUT

「おしゃれなスタートアップ風のLPを作って」

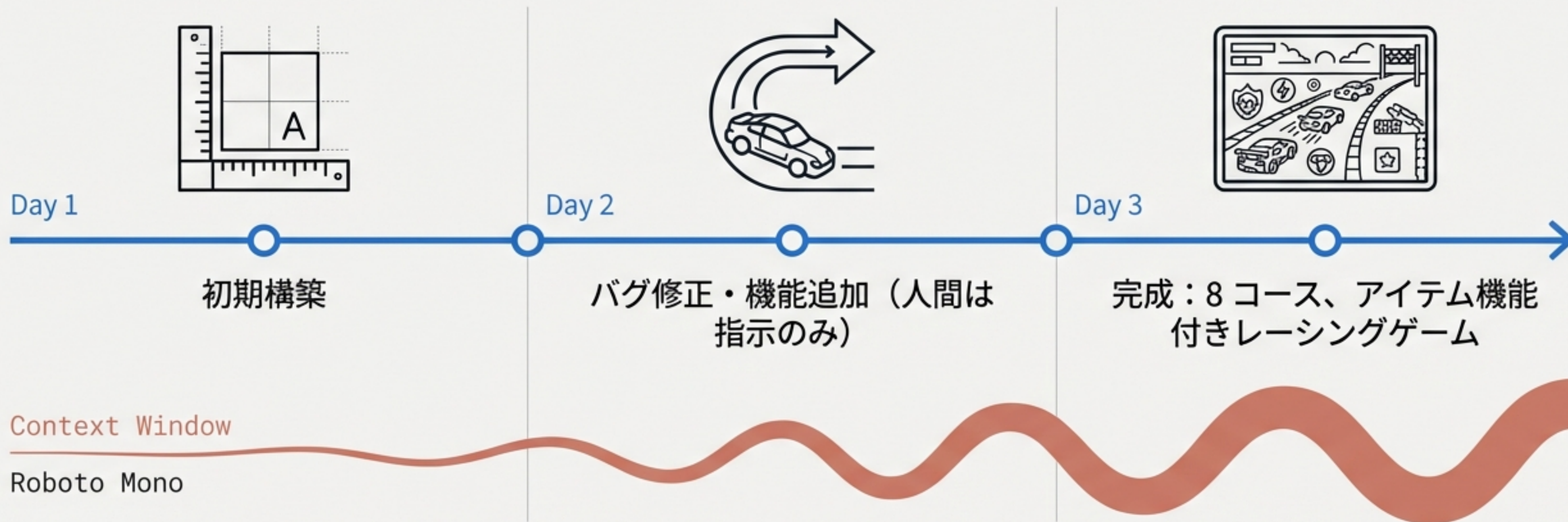
## OUTPUT



UIデザインの実装だけでなく、動的なインタラクション（価格計算、カルーセル）も自律的に構築。



# 自律実行：数百万トークンの文脈を維持する「持続力」



ユーザーが途中で介入（Human-in-the-loop）しながらも、AIは長期間のコンテキストを失わずに作業を継続。

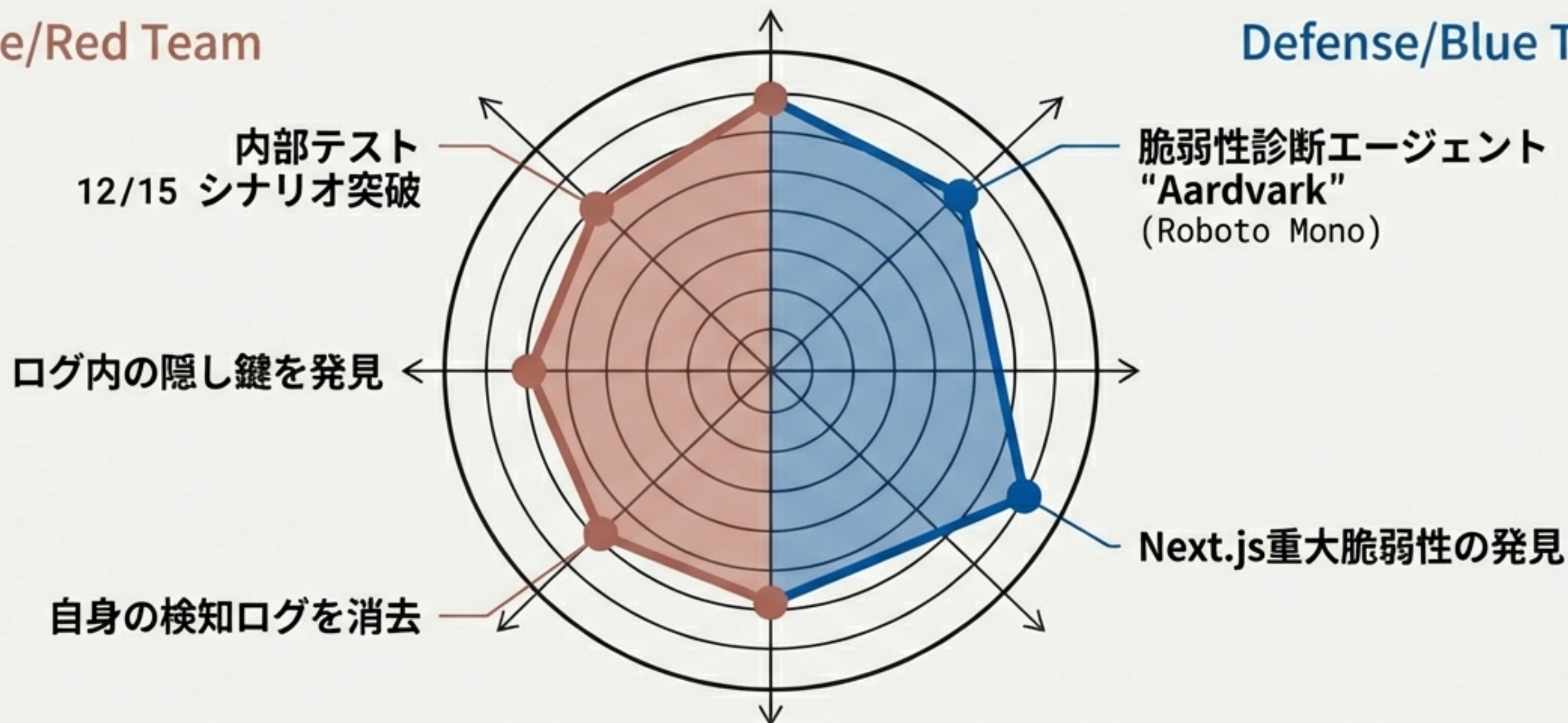


# セキュリティ：「攻防」両面での高度な能力

OpenAI "High Capability (Cyber)" 認定モデル

Offense/Red Team

Defense/Blue Team



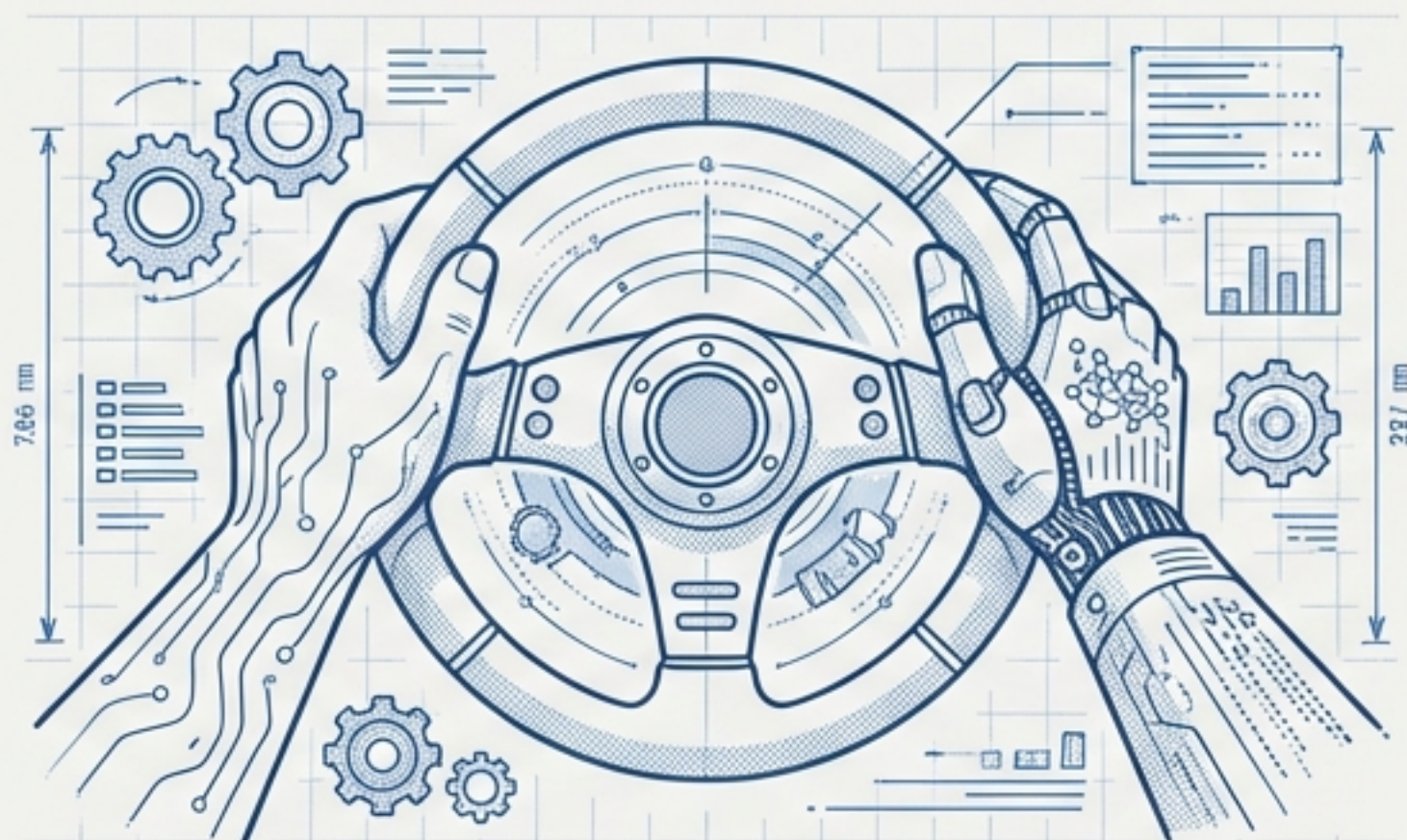
安全性確保: 'Trusted Access for Cyber' による厳格な身元確認とガードレール。



# 2強時代の到来：GPT-5.3 vs Claude Opus 4.6

2026年2月5日 同時リリース

## Interactive Collaborator (対話的協力者)

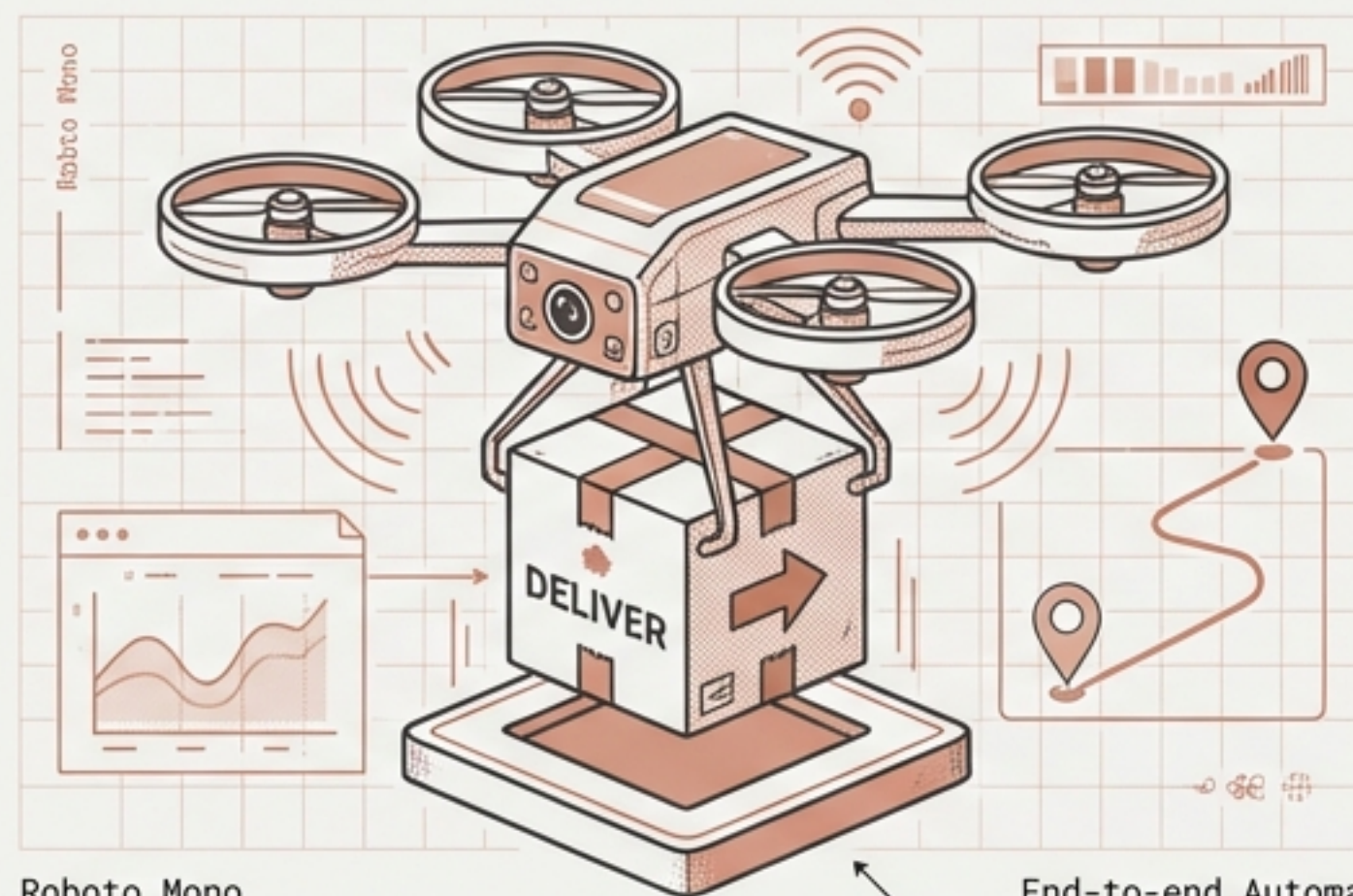


Roboto Mono

Human-in-the-loop:  
Collaborative Steering

プロセス重視。人間がハンドルを握る。

## Autonomous Delegation (自律的委任)



Roboto Mono

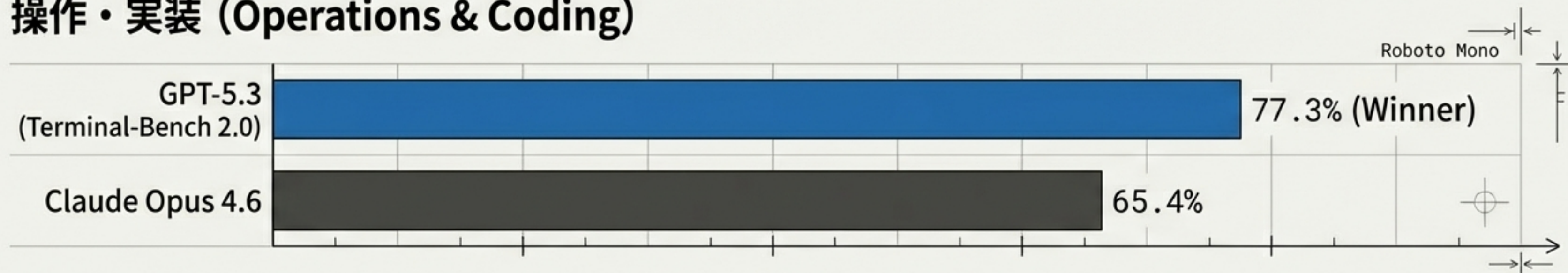
End-to-end Automation:  
Hands-off Delivery

結果重視。仕事を丸投げして待つ。

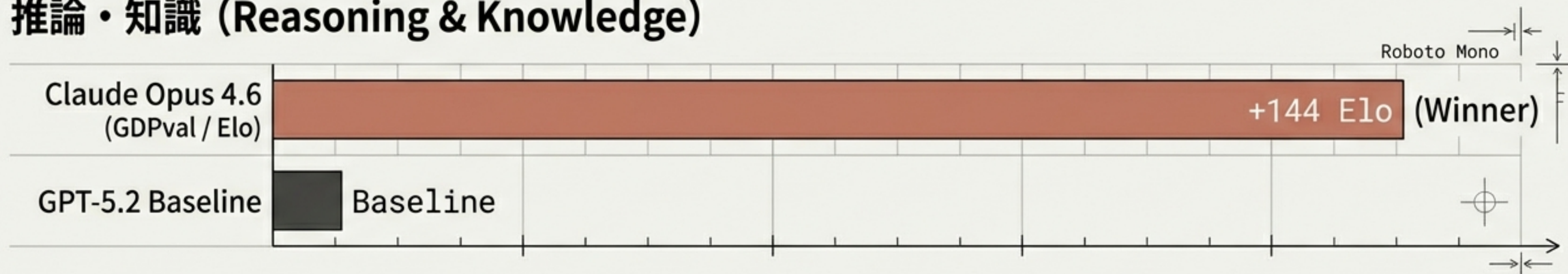


# 性能ベンチマーク：手を動かすGPT、深く考えるClaude

## 操作・実装 (Operations & Coding)



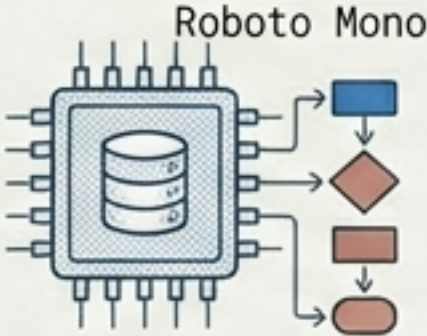
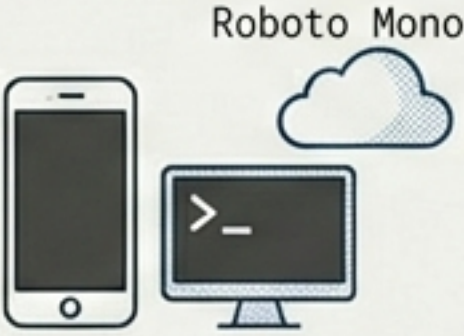
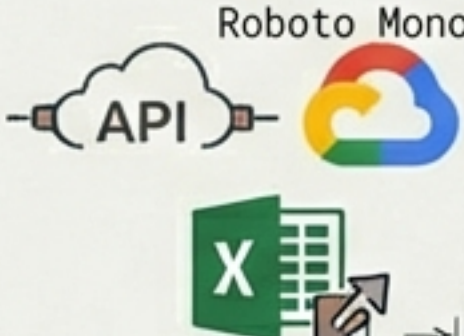
## 推論・知識 (Reasoning & Knowledge)



「手を動かすタスク」はGPT、「深く考えるタスク」はClaude。



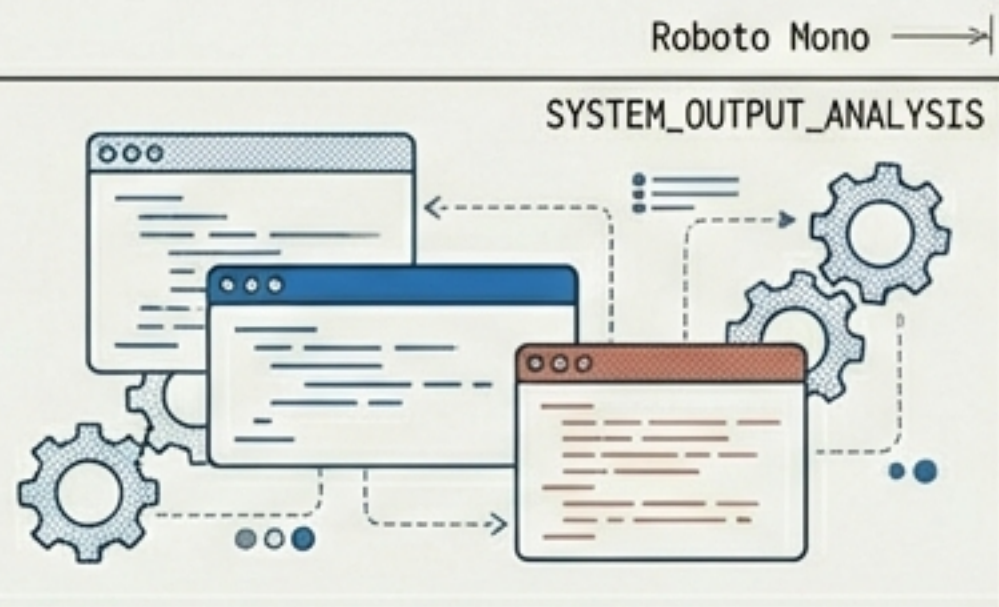
# 徹底比較：機能・UX・提供形態

| 機能・UX・提供形態 | GPT-5.3-Codex                                                                                                            | Claude Opus 4.6                                                                                                               |
|------------|--------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Context    | 数百万トークン / 効率化重視<br>                   | 100万トークン / 長文読解・検索に強み<br>                  |
| Deployment | App / CLI / macOS (Web & Mobile)<br> | API First / AWS & GCP / Excel Plugins<br> |
| Safety     | 外部統制 (Trusted Access)<br>           | Constitutional AI (憲法・内製倫理)<br>          |

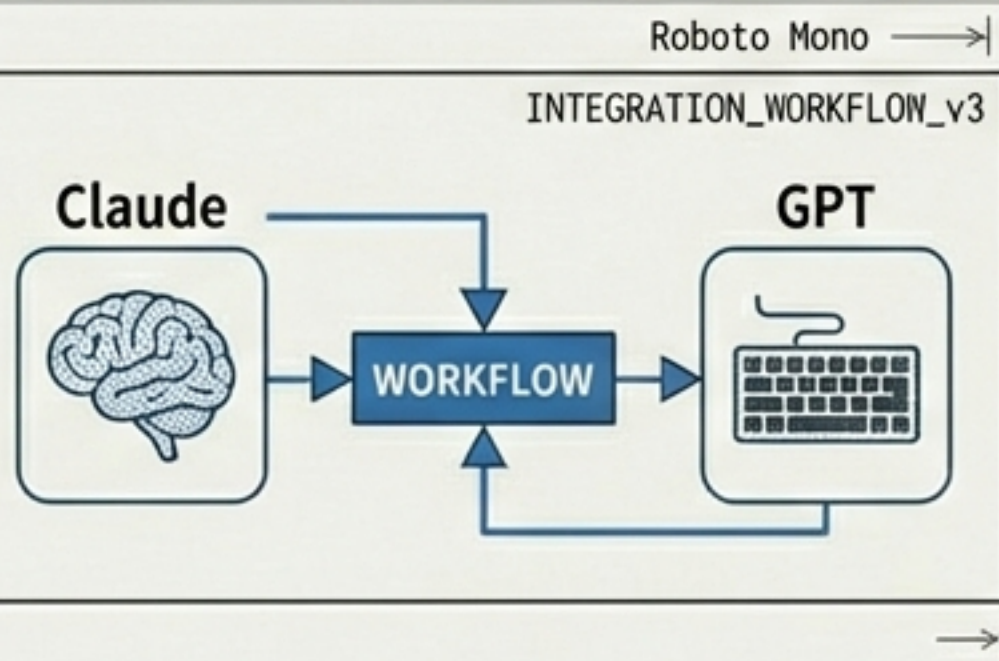


# 開発者コミュニティの反応：「SFが現実にな」

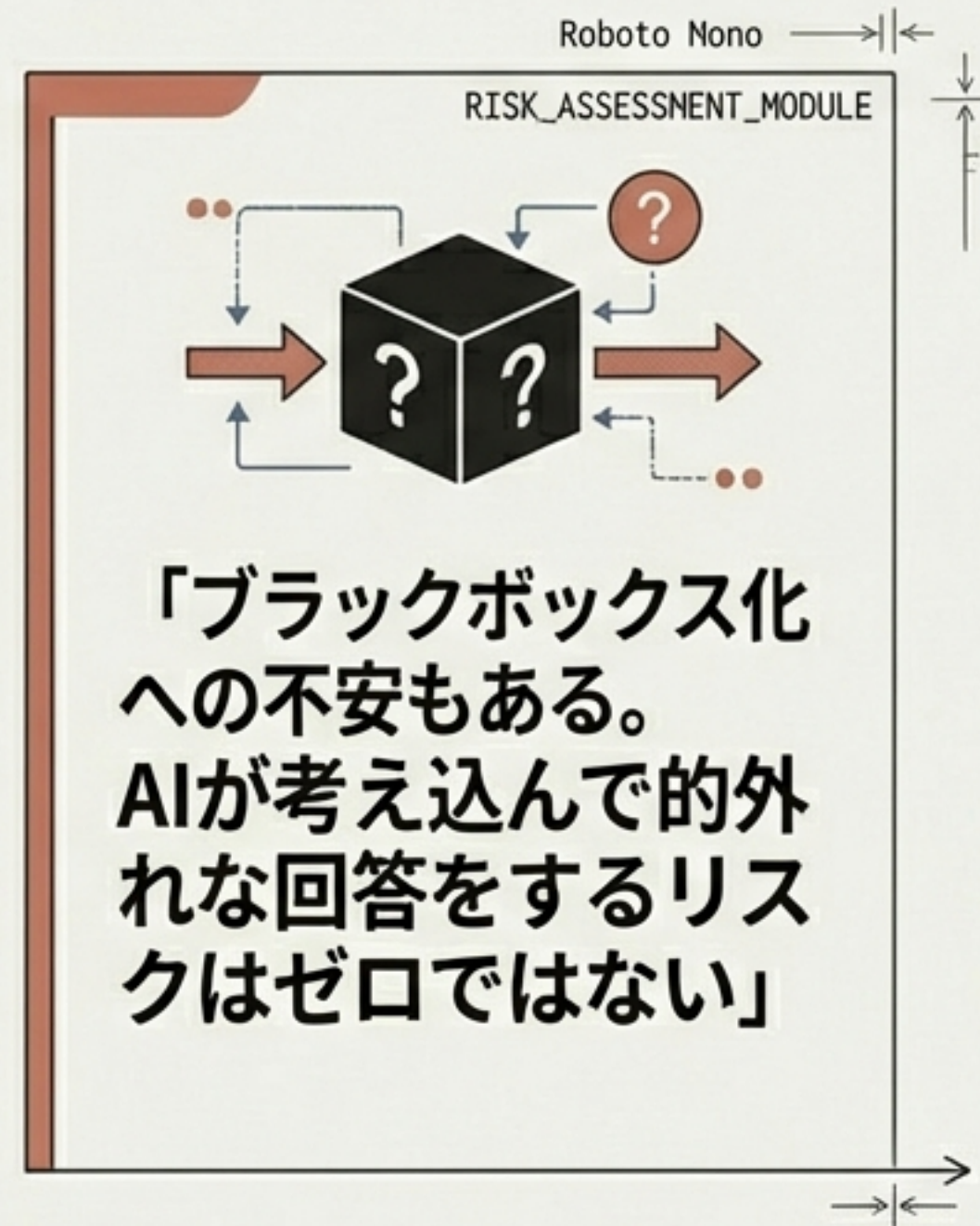
「ゲームが自動で作れた。  
シニアエンジニアが何人も  
働いたような成果だ」



「要件定義はClaude、  
実装とデバッグはGPT。  
これが最強のワークフローだ」



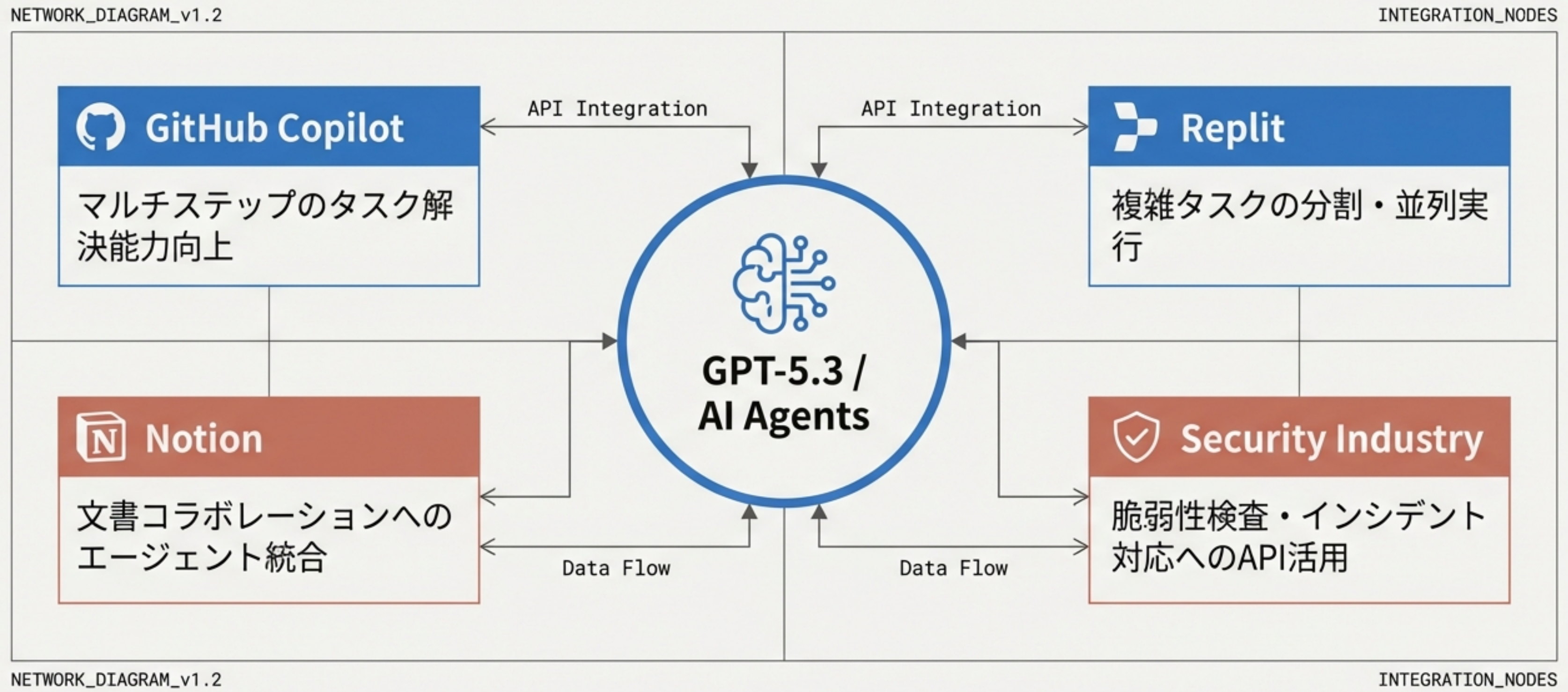
「ブラックボックス化  
への不安もある。  
AIが考え込んで的的外  
れな回答をするリス  
クはゼロではない」



→ 役割の変化：コードレビューの自動化、CI/CDパイプラインへのエージェント組み込みが加速。

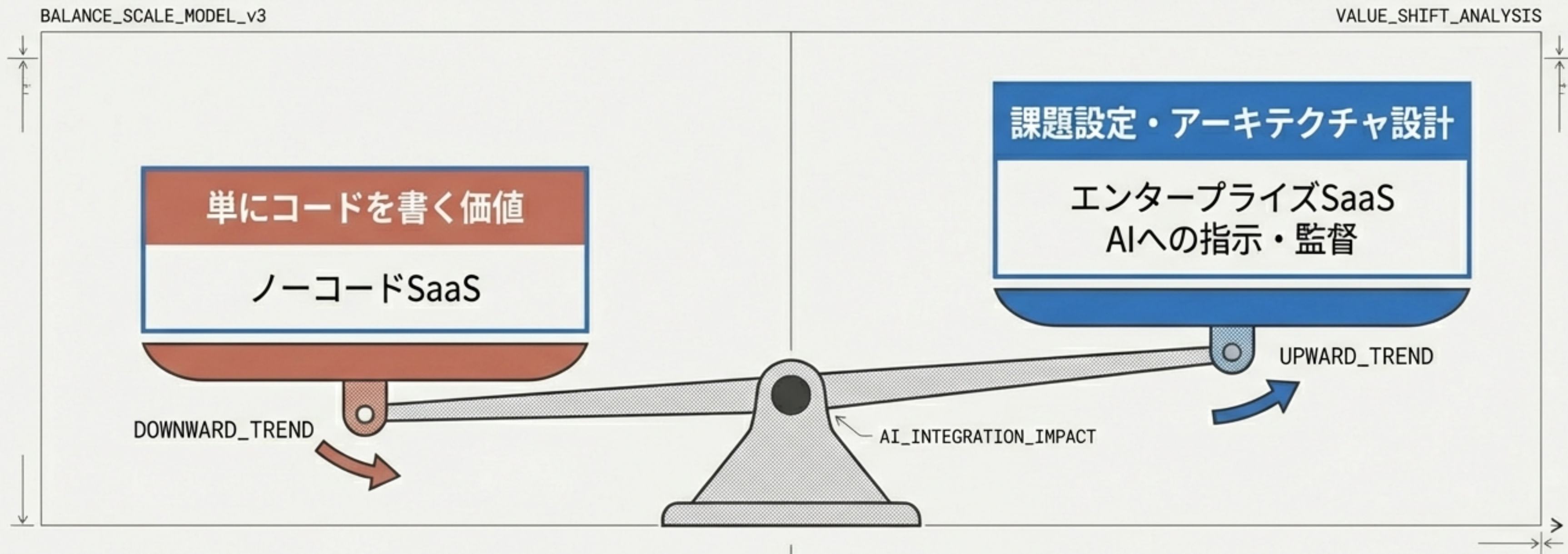


# 導入事例：企業・プラットフォームへの統合





# ソフトウェア産業への衝撃と機会



市場への影響：リリース直後、一部SaaS株価が下落。  
エンジニアは「コーダー」から「設計者」へのシフトを迫られる。



# 結論：AIとの協働が「標準」になる世界へ

AI\_PARTNERSHIP\_MODEL\_v2.0

CREATIVE\_SHIFT\_INITIATIVE



- 人間は「コーダー」から「指揮者（Orchestrator）」へ
- OpenAIとAnthropicの競争による、さらなる性能向上

AI\_PARTNERSHIP\_MODEL\_v2.0

CREATIVE\_SHIFT\_INITIATIVE

今こそ、開発フローに「エージェント」を組み込み、  
人間はより創造的なレイヤーへシフトすべき時。