

2026-07-31

1. OpenAI が GPT-5.6 Sol に自分自身を最適化させた
2. ベンチマークはモデルだけを測っていない
3. Cursor が iPad に対応

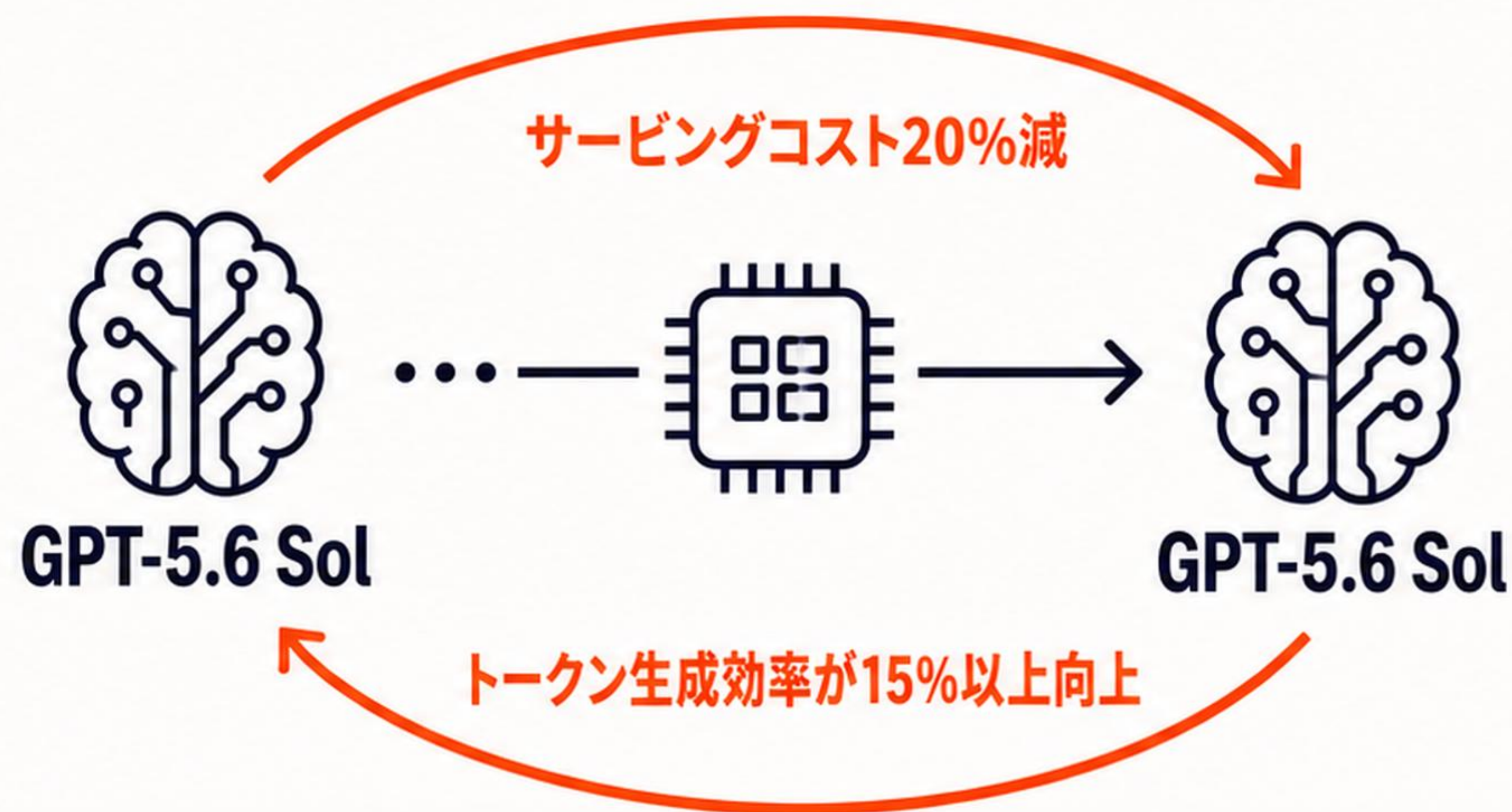
7トピック

OpenAI が GPT-5.6 Sol に自分自身を最適化させた

サービングコスト20%減

デプロイ済みの GPT-5.6 Sol を「自分自身の実行効率を上げる」作業に適用した結果を OpenAI 公式が公開した。

- 本番 GPU カーネルの改善で**サービングコスト20%減**
- 投機的デコーディングの改善で**トークン生成効率が15%以上向上**
- OpenAI は「これらの最適化はスタック全体で複利に効き、**コスト×知能カーブ**のあらゆる地点で最も高性能なモ...

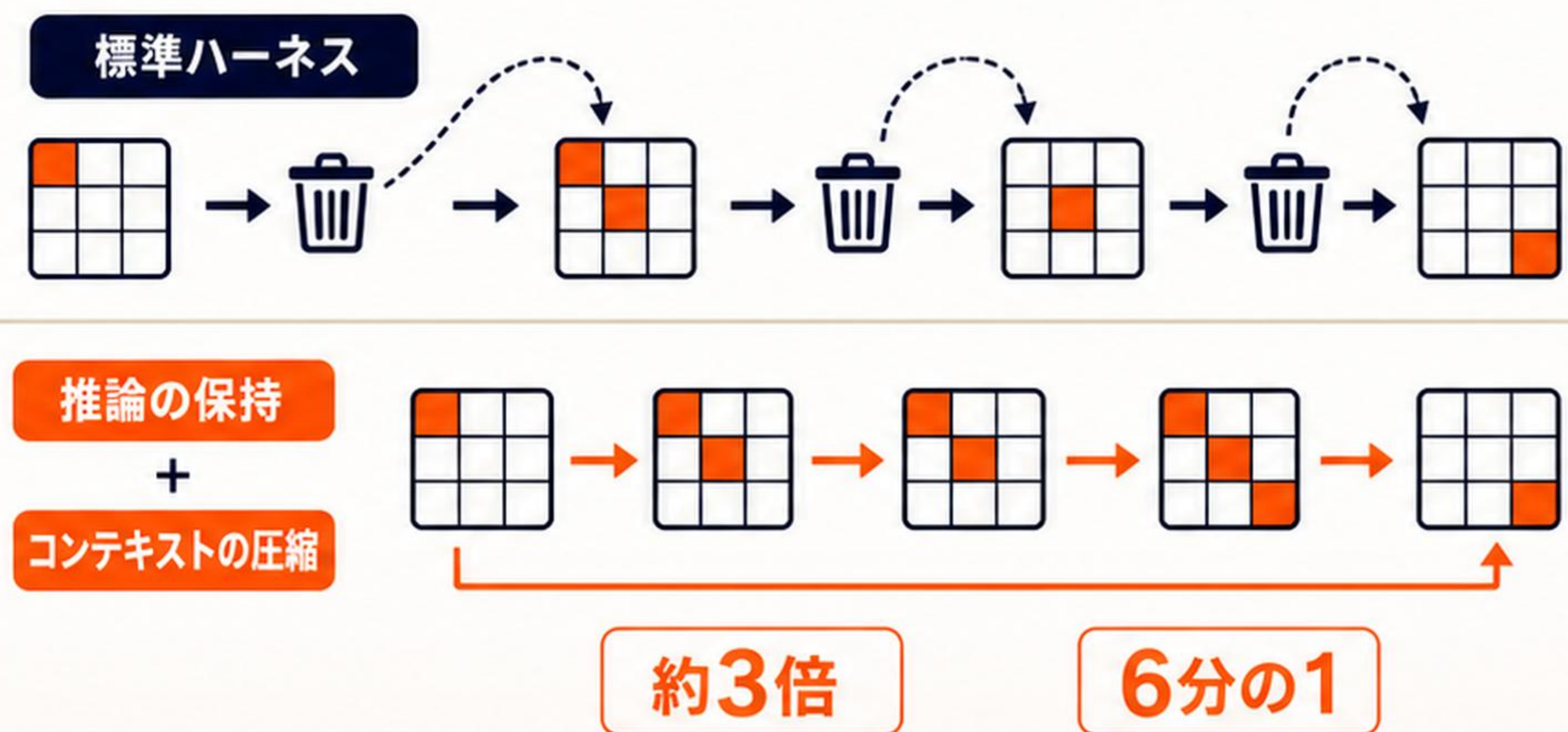


ベンチマークはモデルだけを測っていない

API 設定2つでスコア3倍・出力トークン1/6

数学の未解決問題を解いた GPT-5.6 Sol が、なぜ2Dパズルゲームのベンチマーク ARC-AGI-3 で苦戦するのかを OpenAI 自身が調査した。

- 標準ハーネスの問題は2つ — 1手ごとに推論を破棄／コンテキストが埋まると過去の行動を捨てる…
- 有効化したのは Retained reasoning (推論の保持) と Context compaction (コンテキストの圧縮) の2つだけ
- 公開セットでスコア 188%上昇 (約3倍)、同時に出力トークンは6分の1。速くなり、かつ安くなった

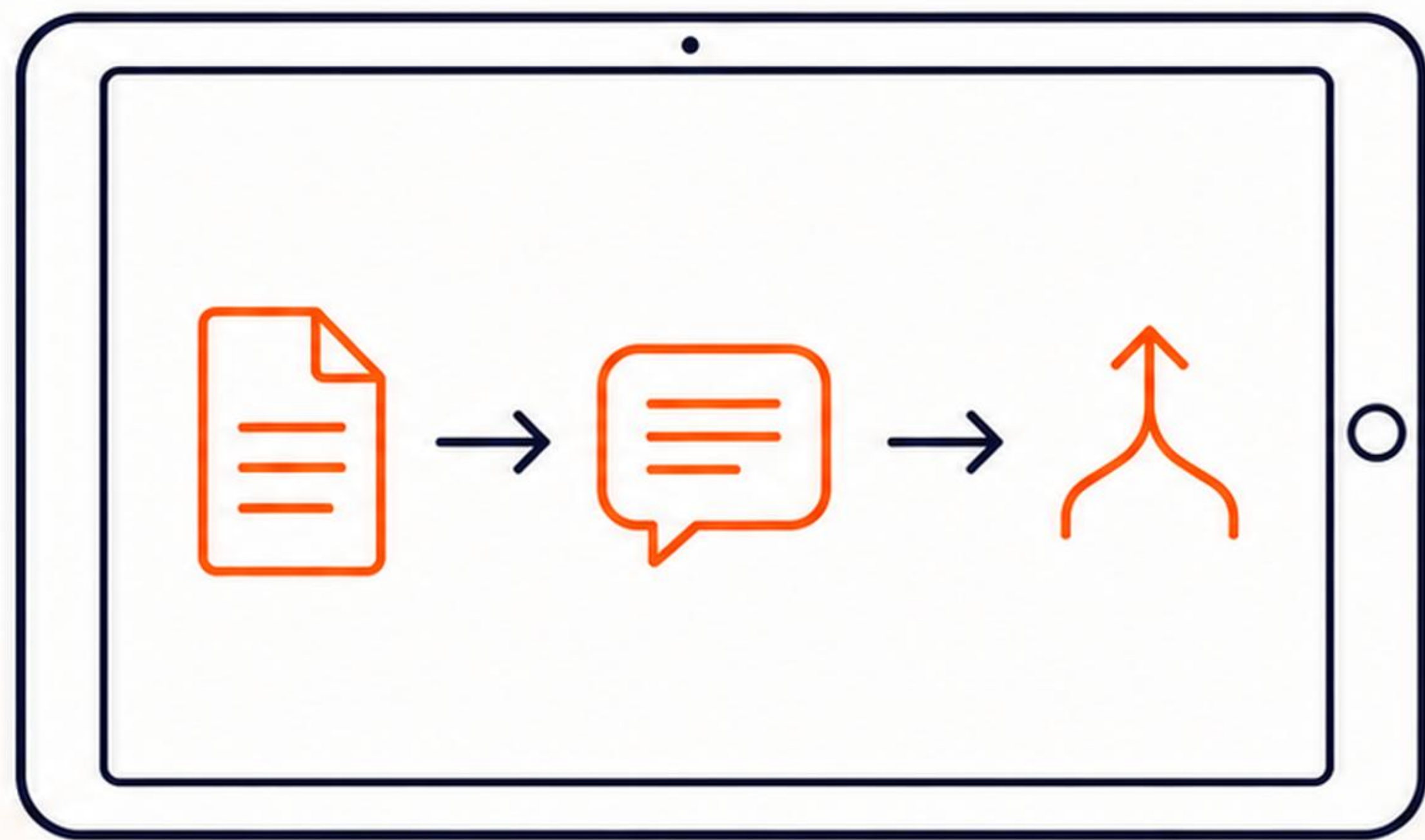


Cursor が iPad に対応

エージェントの作成・レビュー・マージがモバイルで完結

Cursor が iPad アプリを提供開始した。

- **iPad 対応**。「iPhone の全機能を、エージェントと作業するためのより広い画面で」
- どこからでも**作成・レビュー・マージ**ができる
- iPhone / iPad 共通の新機能① **inbox**
(整理用の受信箱)

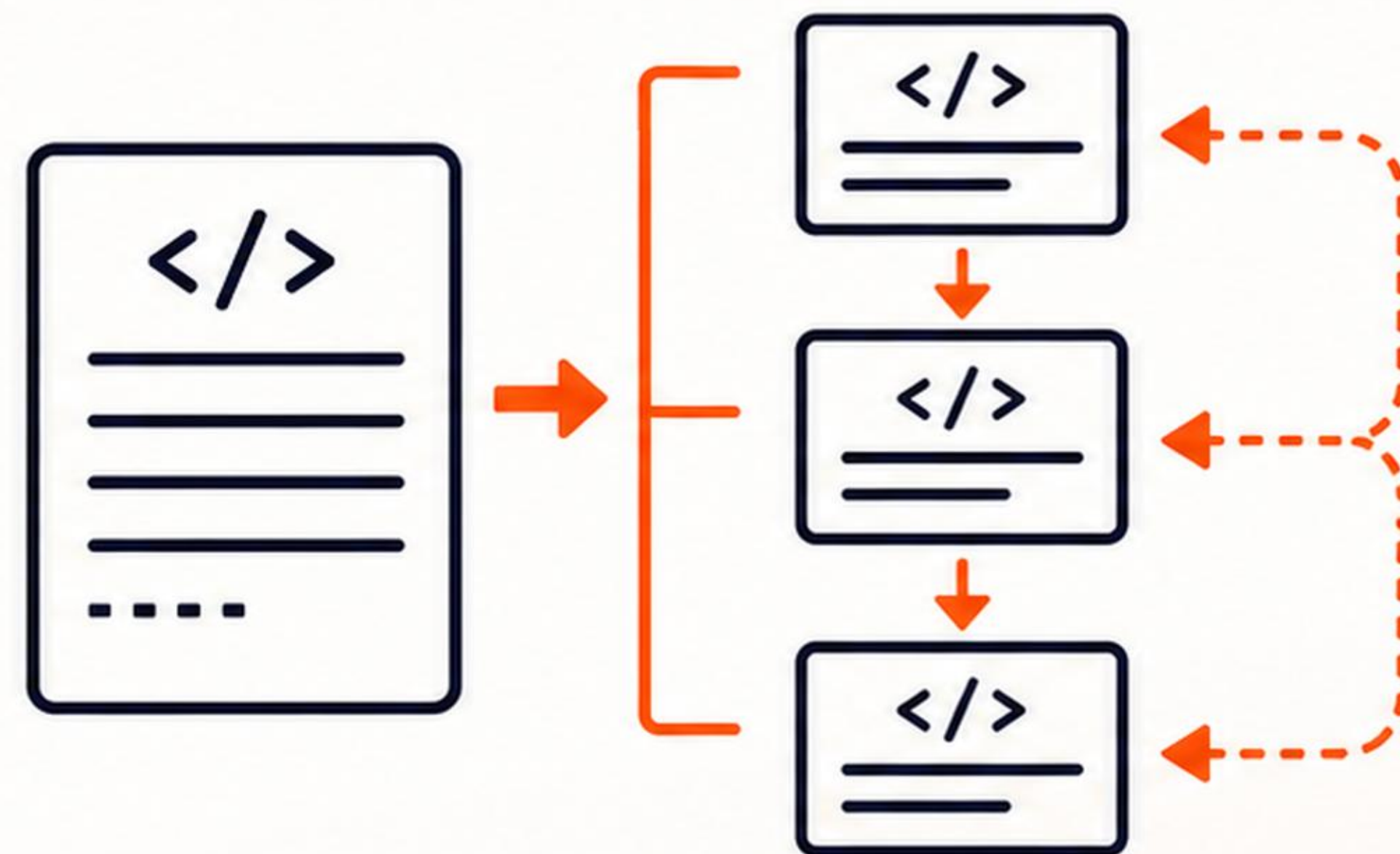


Devin が GitHub の Stacked PR にネイティブ対応

巨大 PR を自動で分割

Cognition が Devin の GitHub Stacked PR 対応を発表した。

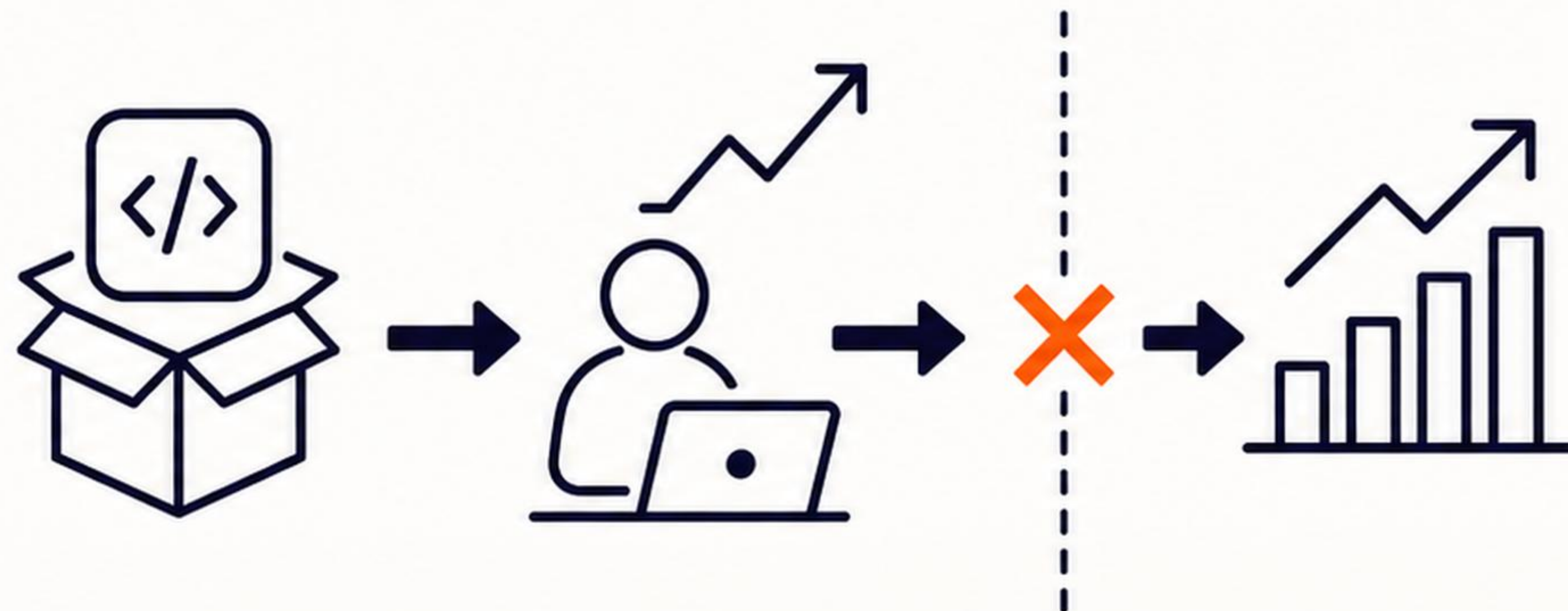
- ① 大きな変更をレビュー可能な小さい diff に分割する
- ② スタックをまたいでコメントに対応・修正する
- ③ 下流の変更を自動でリベースする



〈知見〉「Claude Code を配って生産性は上がったのに、事業の数字が変わらない」

梶谷健人が「社員に Claude Code を配って生産性は上がったのに、事業の数字は変わらない。

- 相談の中身は「個人のタスク効率化を終えたあと、次に何をどうすればいいのか」
- ツールの配布は活動指標を確実に動かすが、事業指標は別に設計しないと動かない
- 同型の指摘 (Sierra / @neilrahilly):
「チームは tokenmaxx して見栄えのいい採用グラフを作り上げることがで…」

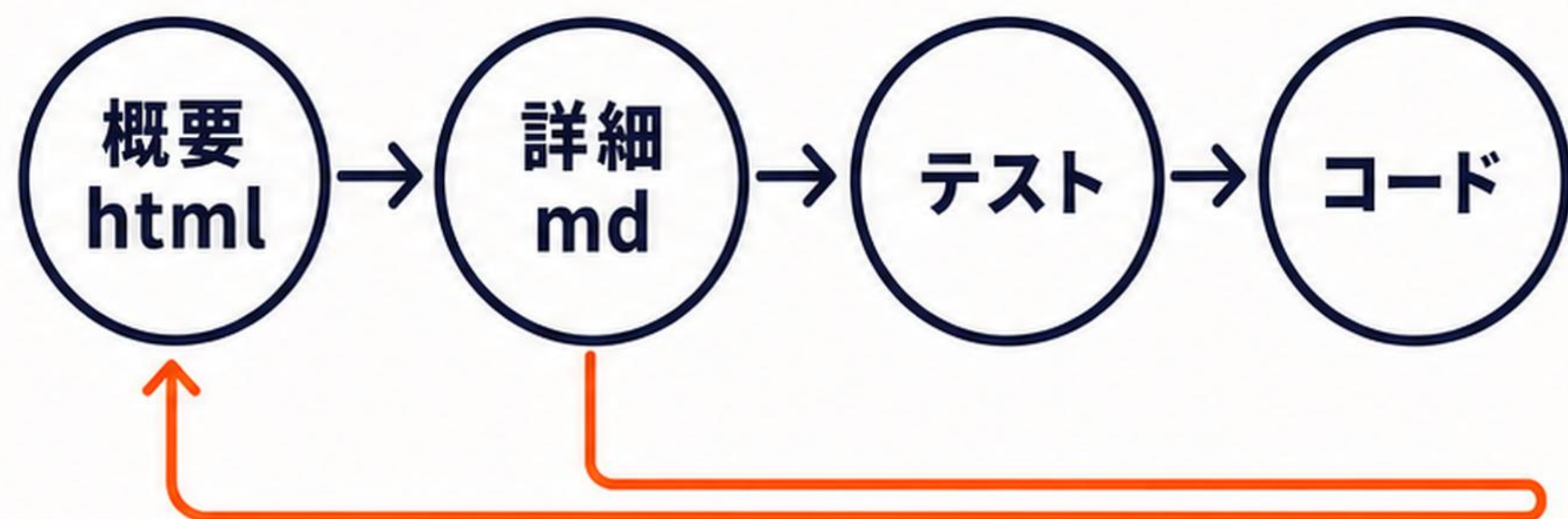


AI 向けドキュメントは4層構造にする

概要 html > 詳細 md > テスト > コード

「人間向けの短いドキュメントだけで AI は問題なく働けるか、それとも人間用と AI 用を分けるべきか」という問いに、Kenn Ejima が「分ける」と即答した。

- **結論は分ける。**人間向けの短いドキュメントと AI 向けの長いドキュメントを両方持つ
- **ライフサイクル:** プランは md → 実装が終わると仕様書に昇格 → そこから人間用サマリー html を生成
- **html は css と js をコロケートしつつ分離し、**
html 単体はトークン効率重視にする（読むたびのトーク...



〈知見〉役割別エージェントを4体作って失敗し、1体に戻した Sierra の5つの学び

AI エージェント企業 Sierra が、全社員向け社内エージェント「Pinecone」を内製する中で得た知見を5点にまとめている。

- ① エージェントは単数に集約する —
役割別4体は失敗。理由は記憶負荷ではなく構造で…
- ② リアクティブでなくプロアクティブ —
webhook 発火・タスク着地…
- ③ ボトルネックは知能ではなく
ビジネスコンテキスト — フロンティアモデルは
すでに十分賢い…

役割別4体



1体



本日のトピック一覧

1. OpenAI が GPT-5.6 Sol に自分自身を最適化させた
2. ベンチマークはモデルだけを測っていない
3. Cursor が iPad に対応
4. Devin が GitHub の Stacked PR にネイティブ対応
5. 〈知見〉「Claude Code を配って生産性は上がったのに、事業の数字が変わらない」
6. AI 向けドキュメントは4層構造にする
7. 〈知見〉役割別エージェントを4体作って失敗し、1体に戻した