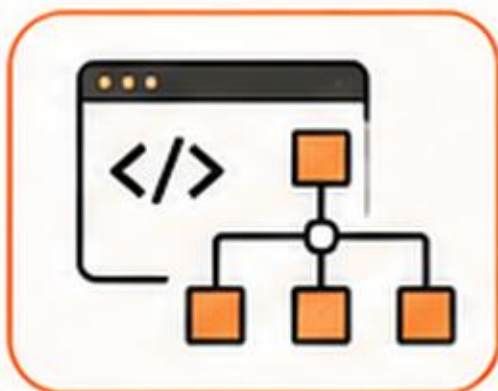




今朝のホットな話題



1



Anthropic「Dynamic Workflows」 — Claude Code が自前の harness を書き、並列サブエージェントを束ねる

2

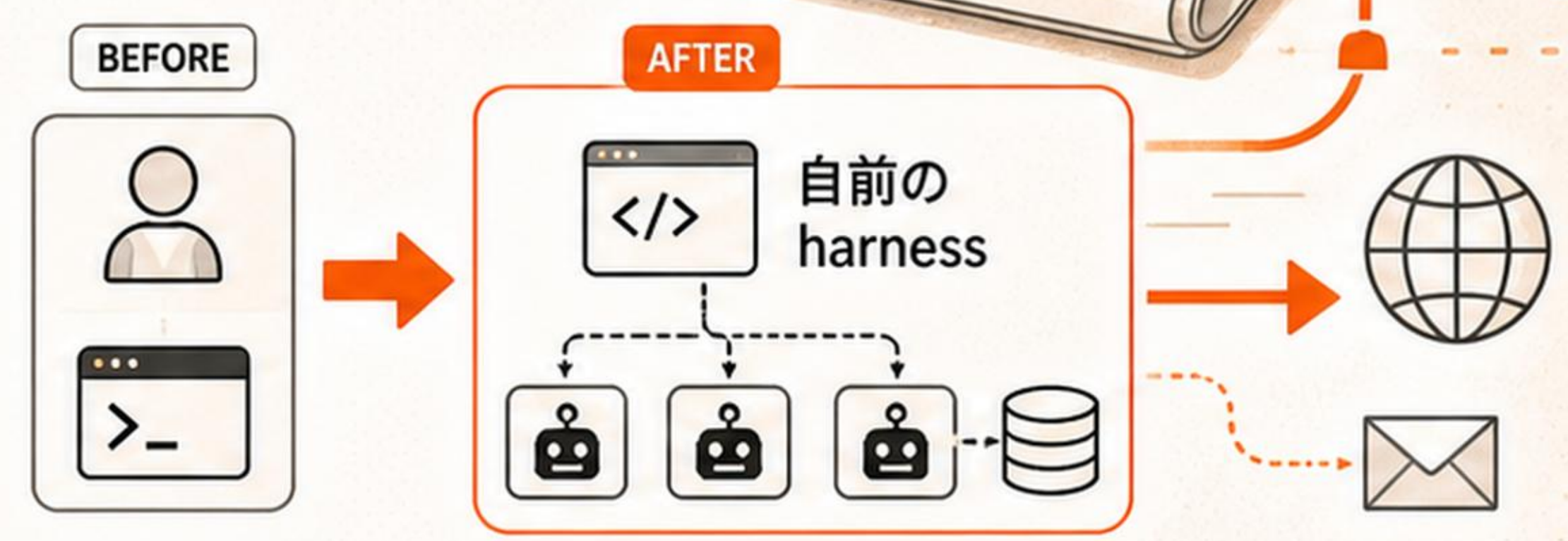


Google「Gemma 4 12B」発表 — エンコーダ不要・16GBメモリのラップトップで動く統合マルチモーダル

3



Anthropic「ant CLI」 — Claude Platform 全体をターミナルから1行で叩ける



7 トピックを整理。



🔍 何が起きた？

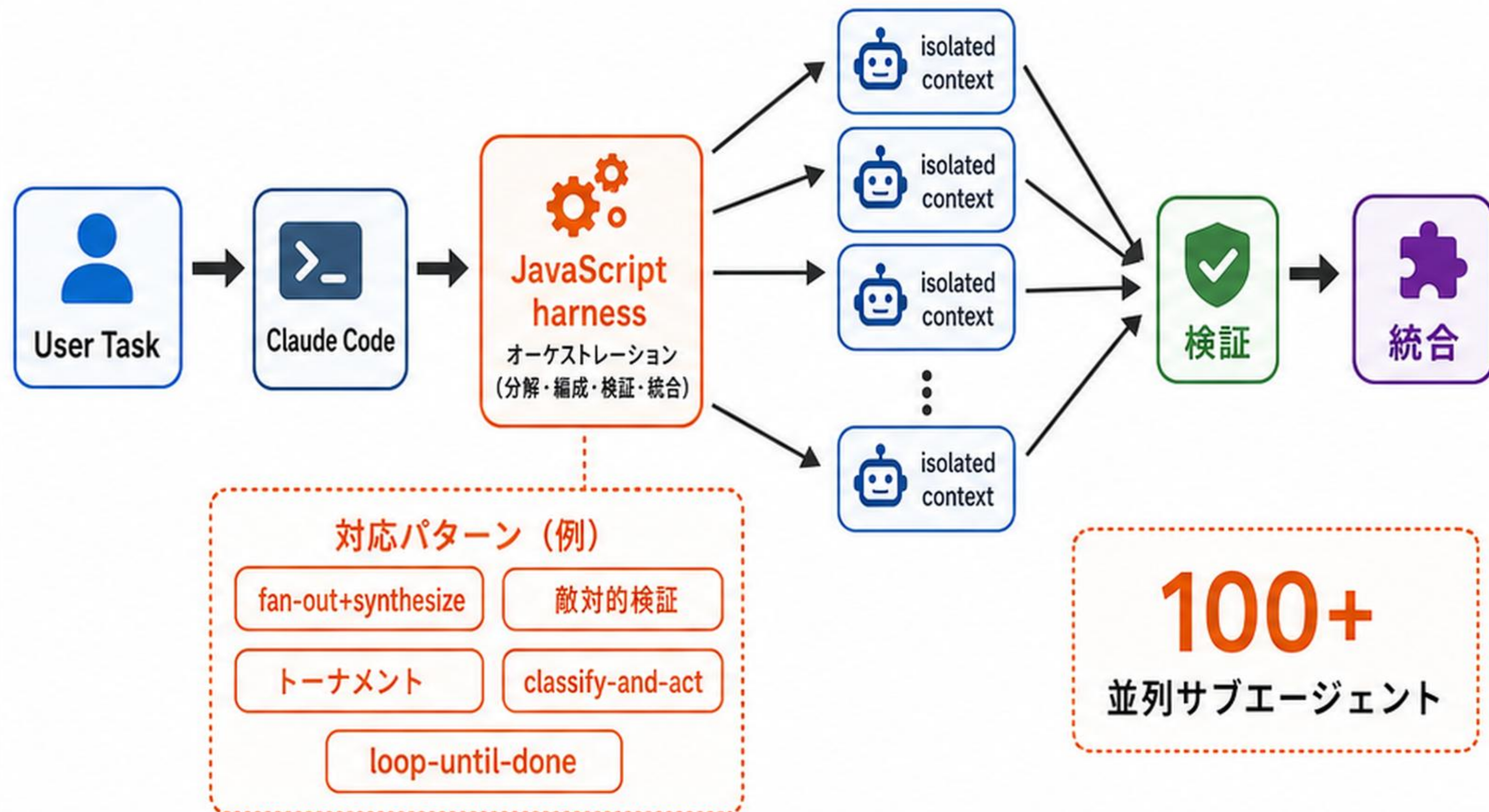
Anthropic が Claude Code に「Dynamic Workflows」を投入。
Claude がタスクごとに独自の JavaScript harness をその場で書き、
複数のサブエージェントを隔離されたコンテキストで起動・検証・
統合する仕組み。

🚩 主な変更点

- Claude が分解・サブエージェント編成・検証・統合をタスクごとに自動設計
- 対応パターン: fan-out+synthesize / 敵対的検証 / トーナメント / classify-and-act / loop-until-done
- 最大100超の並列サブエージェントまでスケール可能と説明
- 用途: ディープリサーチ、大規模リファクタ/マイグレーション、検証、トリアージ、eval

💡 なぜ重要？

単一コンテキストで長時間タスクを回すと生じる手抜き (agentic laziness) ・自己パイアス・目標のズレを、モデルを大きくするのではなくアーキテクチャ側で解く狙い。Opus 4.8 と同時期の打ち手。
トークン消費は増えるが長尺タスクの信頼性が上がる。



@milesdeutscher

Likes 600:

99%が見落とした最強機能



@VibeMarketer_

Likes 446:

harness 論争は終わった



100+ はマーケティング的に盛られがち、という冷静な指摘も

🔍 何が起きた？

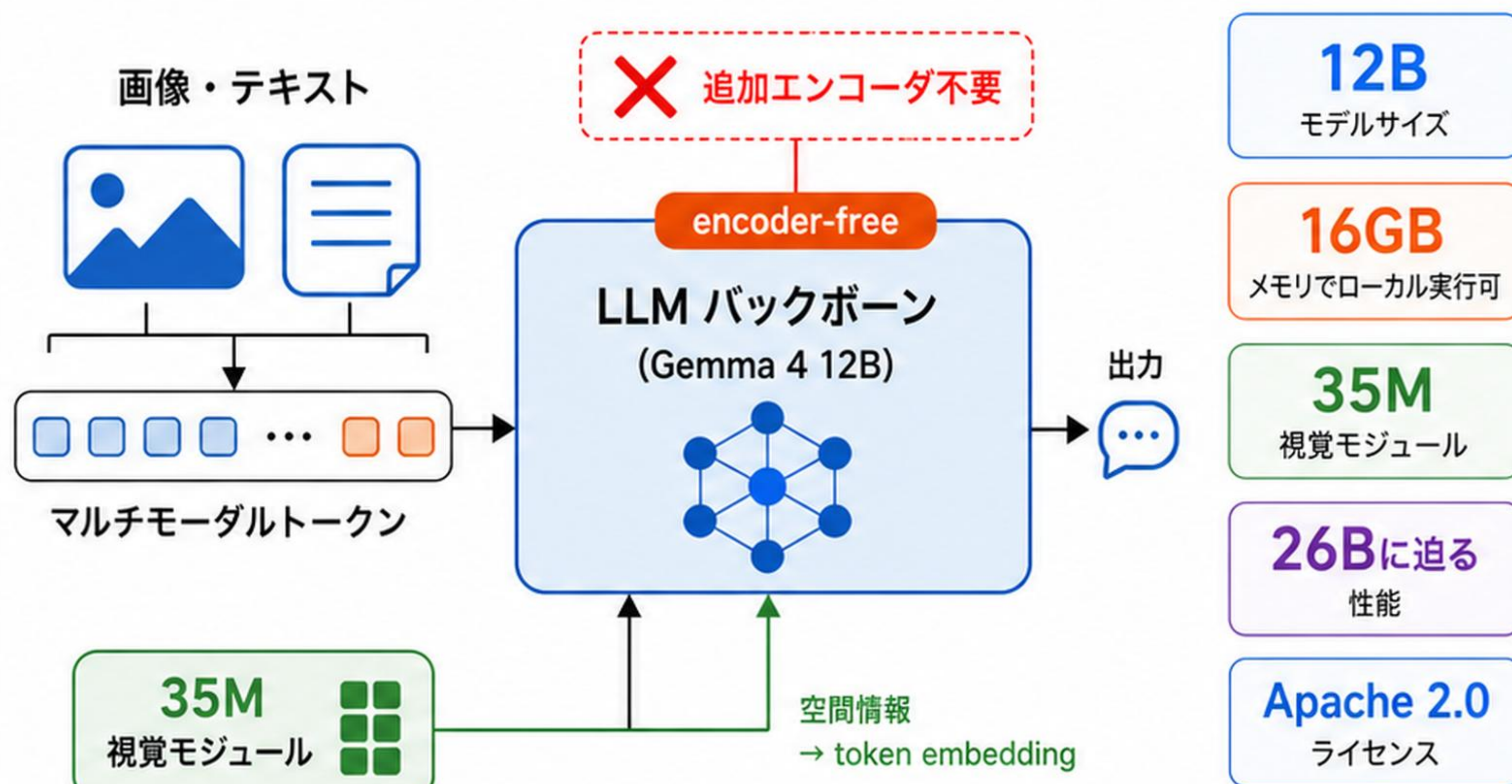
Google が Gemma 4 12B を発表。encoder-free (エンコーダ不要) の統合マルチモーダルアーキテクチャを採用し、画像などのマルチモーダルトークンを LLM バックボーンに直接流す。視覚エンコーダは 35M パラメータの軽量モジュールに置換し、空間情報を token embedding に注入。

📌 主な変更点

- encoder-free: マルチモーダルトークンを LLM バックボーンに直接統合、追加エンコーダ不要
- 視覚エンコーダを 35M パラメータの軽量モジュールに置換
- 16GBメモリでローカル実行可、26Bモデルに迫る性能を半分以下のメモリで
- Apache 2.0 ライセンス、Hugging Face / Kaggle で重み配布
- llama.cpp / MLX / LM Studio / vLLM / SGLang / ollama が初日対応

💡 なぜ重要？

16GB の VRAM/ユニファイドメモリで動き、一回り大きい 26B モデルに迫る性能を半分以下のメモリで実現。ローカルPCで統合マルチモーダルを扱いやすくする。



重み配布

Hugging Face

Kaggle

初日対応プラットフォーム

llama.cpp

MLX

LM Studio

vLLM

SGLang

ollama

🔍 何が起きた？

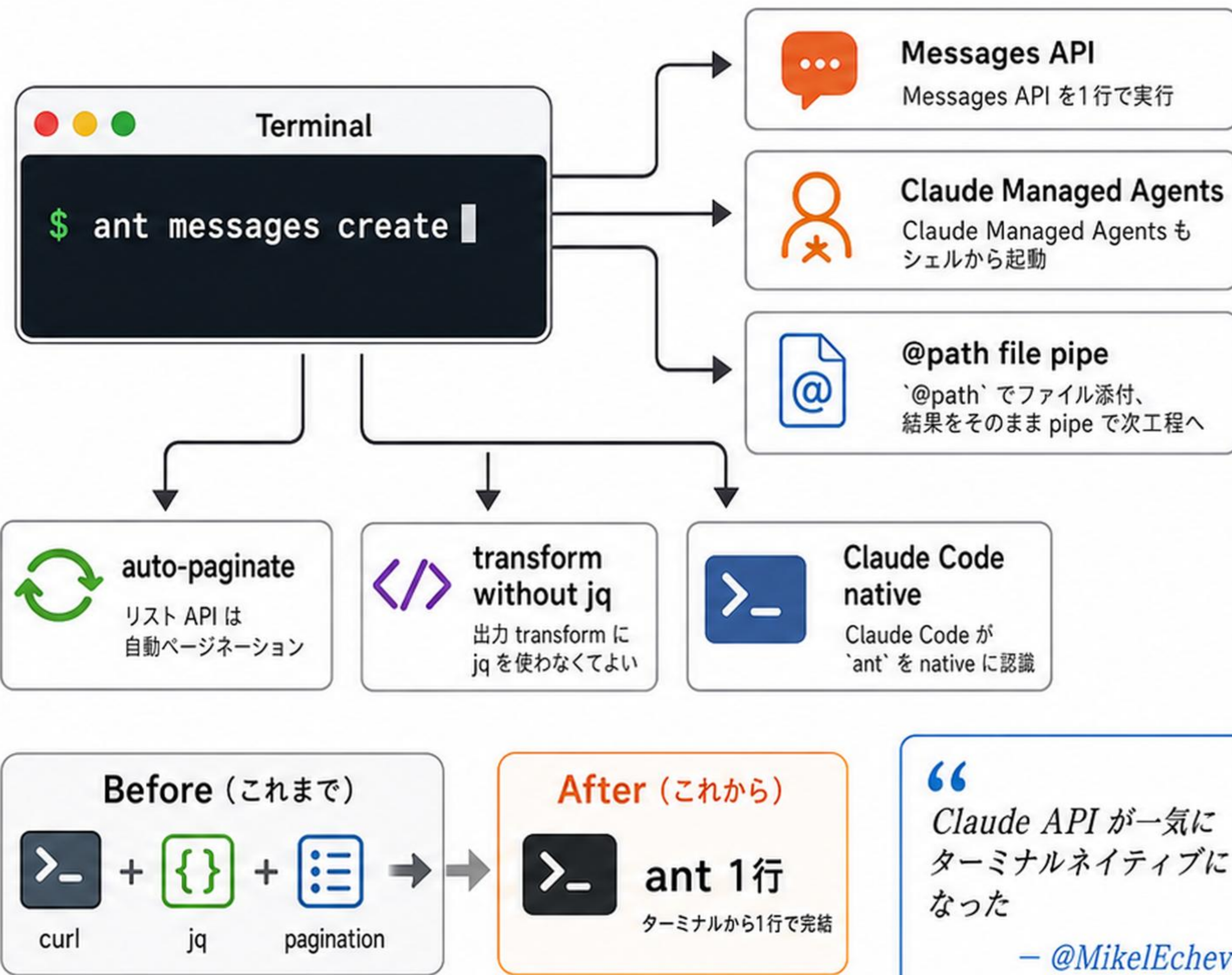
Anthropic が `ant` CLI を公開した。Messages API をはじめ Claude Platform のリソースをシェルから直接呼べるツール。リクエストはフラグまたは YAML で組み立て、`@path` でファイルを pipe、リスト系 API は自動ページネーション、出力整形は jq 不要で完結する。Claude Code がネイティブに認識する。

📌 主な変更点

- Messages API を1行で実行、Claude Managed Agents もシェルから起動
- `@path` でファイル添付、結果をそのまま pipe で次工程へ
- リスト API は auto-paginate、出力 transform に jq を使わなくてよい
- Claude Code が `ant` を native に認識

💡 なぜ重要？

Claude API が一気にターミナルネイティブになり、dev フレンドリーさが増した。Xでは @MikelEcheve が紹介。



🔍 何が起きた？

OpenAI が Codex の新機能「Sites」を発表。Codex アプリ内で作った Web サイト/アプリをワークスペース内に公開・URL共有できる。

📌 主な変更点

- Sites: Codexアプリ内でWebサイト/アプリをワークスペースに公開・URL共有
- 従来 Cloudflare D1/Turso+Workers で安価公開できた手段をラップして誰でも使えるように
- Lovable / Replit / Wix がローンチパートナーかつ競合という構図

💡 なぜ重要？

@sarukun99 は「ローカルで作る段階は既の実現済みで、ネックだった本番公開のピースを Sites が埋めた。ネットワーク効果を必要としない大半の SaaS は1時間で作って公開できる」と主張し「SaaSは死んだ」論争を呼んだ。

個別最適で作るコストが共有SaaSより安くなり、上流工程（業務整理）の価値が上がる。

✂️ xでの反応

「SaaSは死んだ」に賛否。@Tsunoda_LegalOn は「既存SaaSを全て殺すという意味なら誤り」と慎重論。



🔦 要点

AI に書かせたコードを安全に本番運用するための設計論。
『ソースコードが完璧』前提を捨て、コードは不完全である前提で、その周囲を境界（サンドボックス）で囲って守る。

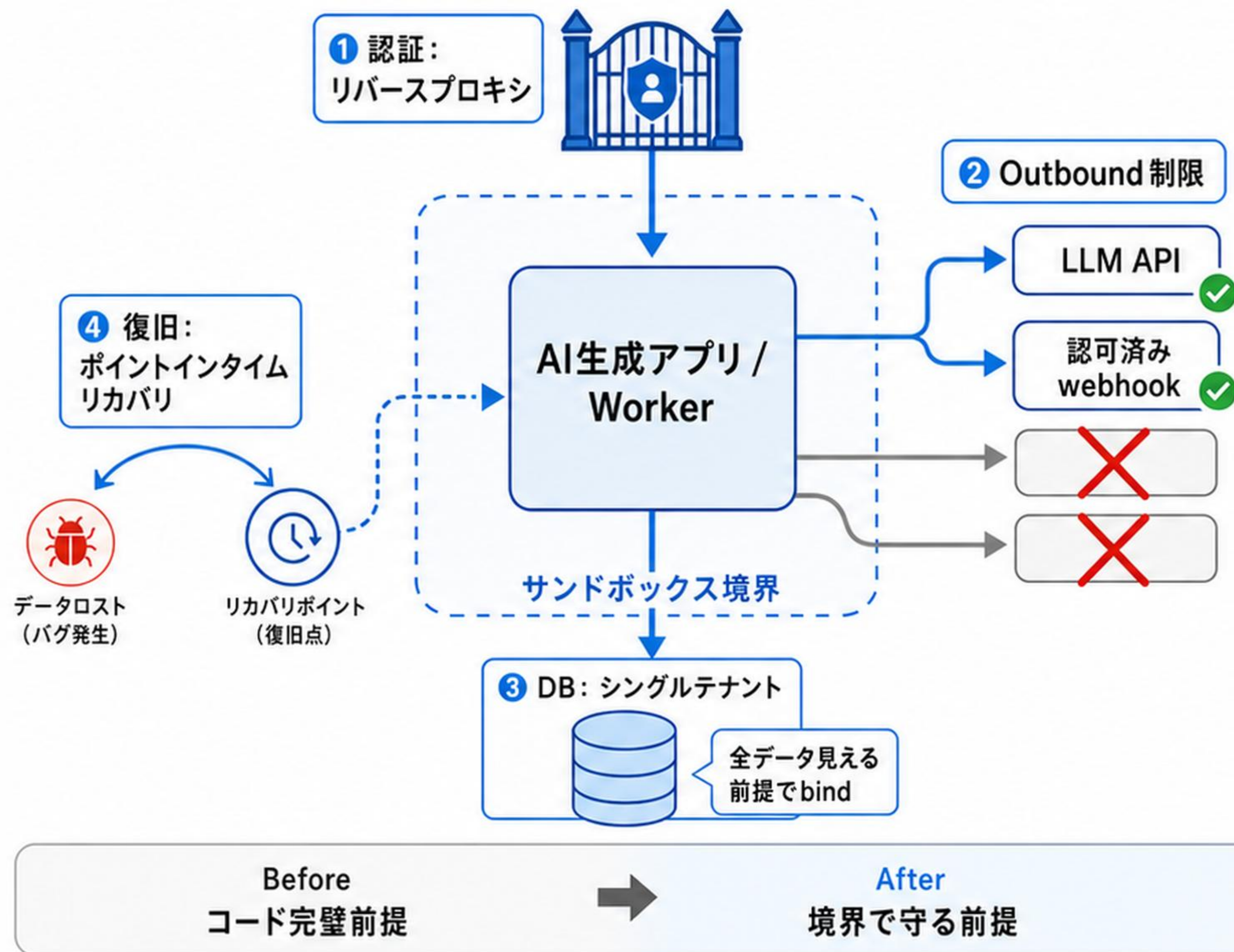
🔧 具体的な手法 / 使いどころ

- 認証はリバースプロキシで外出し（ALBのGoogleログイン済みチェックと同様）、配下アプリは認証ロジック不要
- outbound を絞り、LLM API か認可済み webhook 以外の送信先を塞ぐ
- DBをマルチテナントからシングルテナント化し『全データ見える前提』でWorkerにbind
- プログラムのバグによるデータロストはポイントインタイムリカバリで対応

🌱 なぜ刺さるか / 学び

不完全なコードを前提に、防御面を単純化する。境界で囲い、外部送信・認証・DB分離・復旧を設計で担保する。

不完全なコードを境界で囲う



🔍 何が起きた？

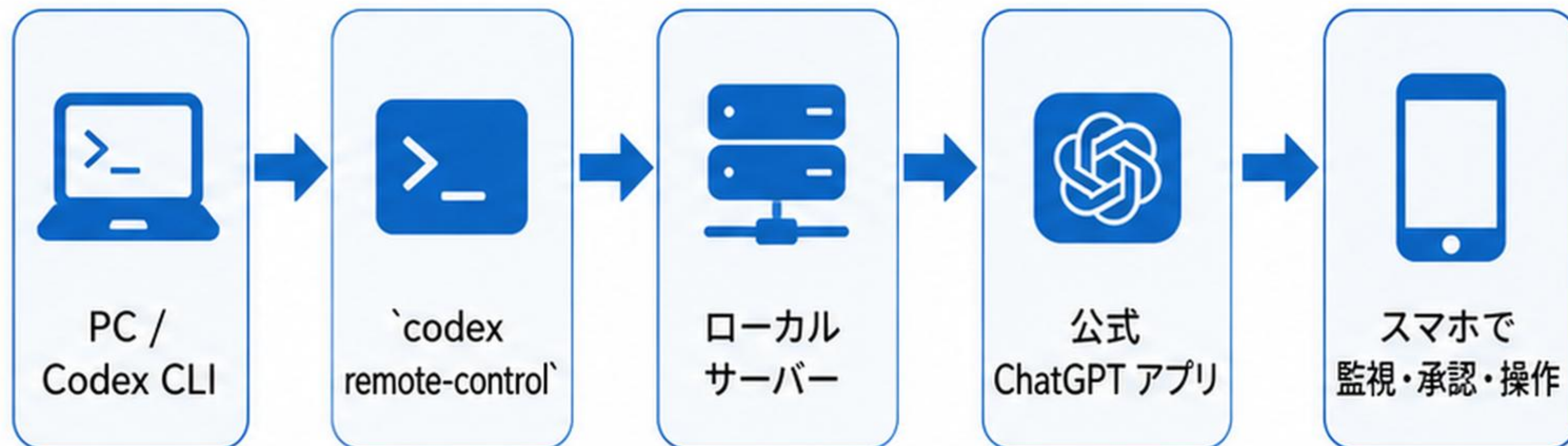
OpenAI が Codex を ChatGPT モバイルへ統合する流れの中で、`codex remote-control` でローカルサーバーを起動すれば公式 ChatGPT アプリからスマホ経由で Codex CLI を操作できることが確認された。PCで動くAIエージェントを移動中にスマホから監視・承認・操作でき、サードパーティ製の遠隔ツールが不要になる。Anthropic も Claude Code で同様の遠隔機能を提供済み。

📌 主な変更点

- `codex remote-control` でローカルサーバー起動 → 公式ChatGPTアプリから操作（実証あり）
- 移動中もエージェント開発を監視・承認・継続できる
- 「AIが自律で働き、人間が指揮する」遠隔指揮スタイルが標準化の兆し
- ChatGPT×Codex のアプリ統合（スーパーアプリ化）は community で議論先行・一部は憶測

💡 なぜ重要？

公式アプリ経由の遠隔操作により、開発者はPC前にいなくてもAIエージェントの作業を見守り、承認し、継続できる。遠隔指揮型の開発体験が標準UIになりつつある。



Xでの反応: @kazuyamano
「有料の遠隔ツールに課金したが公式 remote-control で十分だった」



Anthropic / Claude Code も同様の遠隔機能

🔍 何が起きた？

GPT-5.6 が近日中にリリースされるとの観測が広がっている。GPT-5.4(3/5)→5.5(4/23)→5.6(6/4?) と約6週間間隔でメジャーモデルが出るパターンが注目されている。

📌 主な変更点

- モデルカード・ベンチマーク・正式日は未公開 (名前のみCodexログで観測)
- リリースサイクルが約60日に短縮しているという指摘
- Polymarket で6月末までのリリースに高確率 (80~89%)

💡 なぜ重要？

状況証拠はあるが未確定。期待先行で『スーパーアプリ同時発表では』との憶測も出ているため、正式なモデルカード・ベンチマーク・リリース日の確認が重要。Xでは @daniel_mac8 (Likes 251) が規則性を指摘。



⚠️
未確定



Codexログ:
名前を観測



Polymarket:
6/30まで
80~89%



未公開:
モデルカード /
ベンチマーク /
正式日

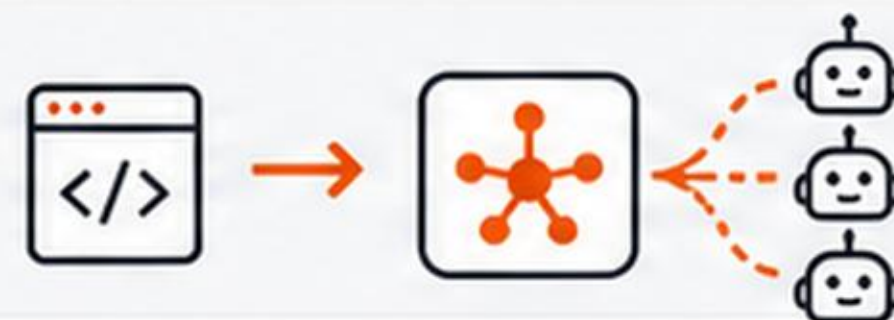


@daniel_mac8: リリースサイクルの規則性を指摘

📍 Likes 251

本日のトピック一覧

1. Anthropic 『Dynamic Workflows』 — Claude Code が自前の harness を書き、並列サブエージェントを束ねる



2. Google 『Gemma 4 12B』 発表 — エンコーダ不要・16GBメモリのラップトップで動く統合マルチモーダル



3. Anthropic 『ant CLI』 — Claude Platform 全体をターミナルから1行で叩ける



4. OpenAI 『Codex Sites』 — ローカル開発から本番公開まで1時間、「SaaSは死んだ」論争



5. バイブコーディング時代のセキュリティ — 「不完全なコードを境界で囲う」防御論



6. Codex を公式 ChatGPT アプリ+スマホから遠隔操作 — codex remote-control



7. GPT-5.6 リリース間近の観測 — OpenAI の約60日リリースサイクル (未確定)

