Tech News Daily

MORNING DISPATCH

2026-08-11

1. Anthropic、未公開の研究版 Claude がリーマンゼータの下界を 41.6% → 67.2% に更新
2. OpenAI、サイバーセキュリティ専用モデル GPT-5.6-Cyber を投入し Daybreak を拡張
3. Anthropic、Claude Sonnet 5 の導入価格 $2/$10 を恒久化（9/1 からの値上げを撤回）

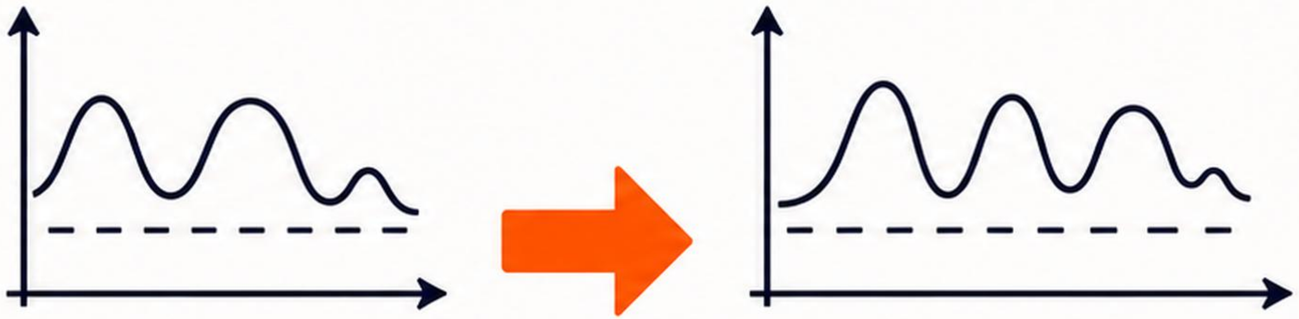
8トピック

TEKION Group

Anthropic、未公開の研究版 Claude がリーマンゼータの下界を 41.6% → 67.2% に更新

Anthropic が、未公開の研究版 Claude にリーマン予想へ挑ませた試みの結果を公開した。

- 下界の更新は 41.6% → 67.2%。従来の 41.6% は数学者たちが数十年かけて到達していた記録
- 作業は Claude Code 上の2セッション、出力トークン合計3,100万
- 約60個のサブエージェントを動かし、2,400のシェルコマンドを実行、数百のPythonスクリプトを書いた



「AIが数学を解いた」ではなく、未解決問題に向かって60個のサブエージェントを3,100万トークン分ぶつける運用が、実際に記録を更新したことのほうが読みどころ。

OpenAI、サイバーセキュリティ専用モデル GPT-5.6-Cyber を投入し Daybreak を拡張

OpenAI がサイバーセキュリティ施策 Daybreak を拡張し、認可された高度なセキュリティ業務向けの新モデル GPT-5.6-Cyber を投入した。

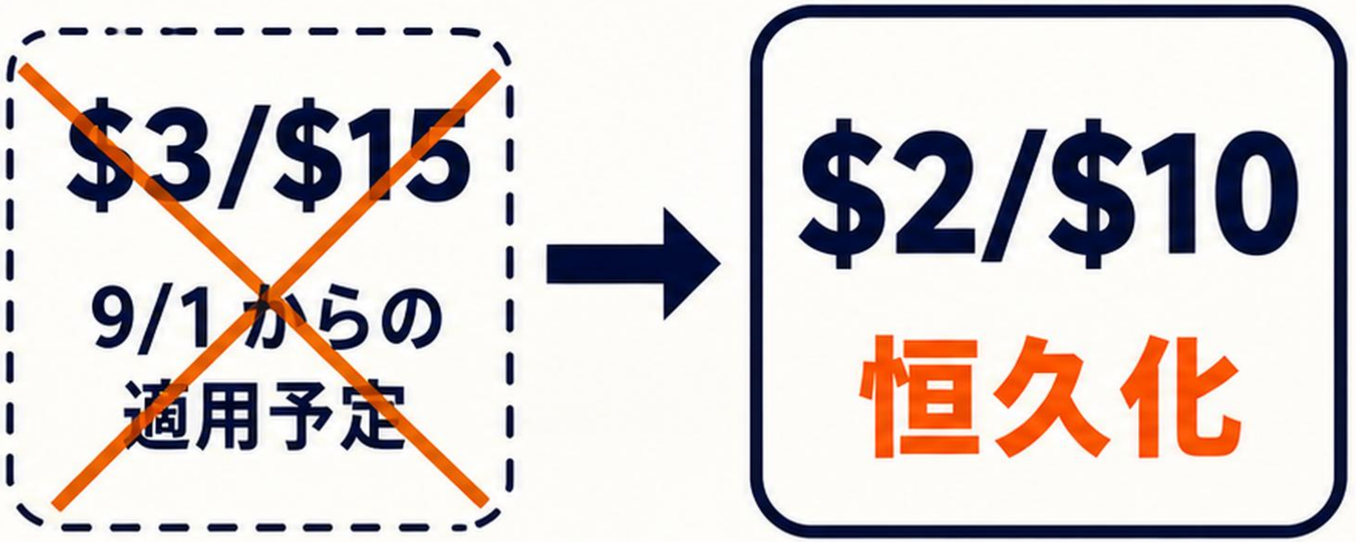
- **Daybreak Blue:** GPT-5.6 Sol を含むフロンティアモデルに、広い防御業務向けにキャリブレーションされたセ…
- **Daybreak Red:** GPT-5.6-Cyber を含む専用訓練モデルへのアクセス。認可された脆弱性研究…
- アクセスは**承認された防御側に限定**され、リスクの高い業務には追加の制御と監視が付く

汎用モデルの安全性を上げる方向ではなく、能力を上げたうえで配布先を絞るという設計判断。

Anthropic、Claude Sonnet 5 の導入価格 $2/$10 を恒久化（9/1 からの値上げを撤回）

Anthropic が Claude Sonnet 5 の導入価格を恒久化すると発表した。

- 恒久化されるのは 入力 **$2** / 出力 **$10**（100万トークンあたり）
- 予定されていた標準価格は **$3/$15**。9/1 からの適用が撤回された
- Sonnet 5 のローンチは 2026年6月30日、導入価格の期限は 8/31 だった



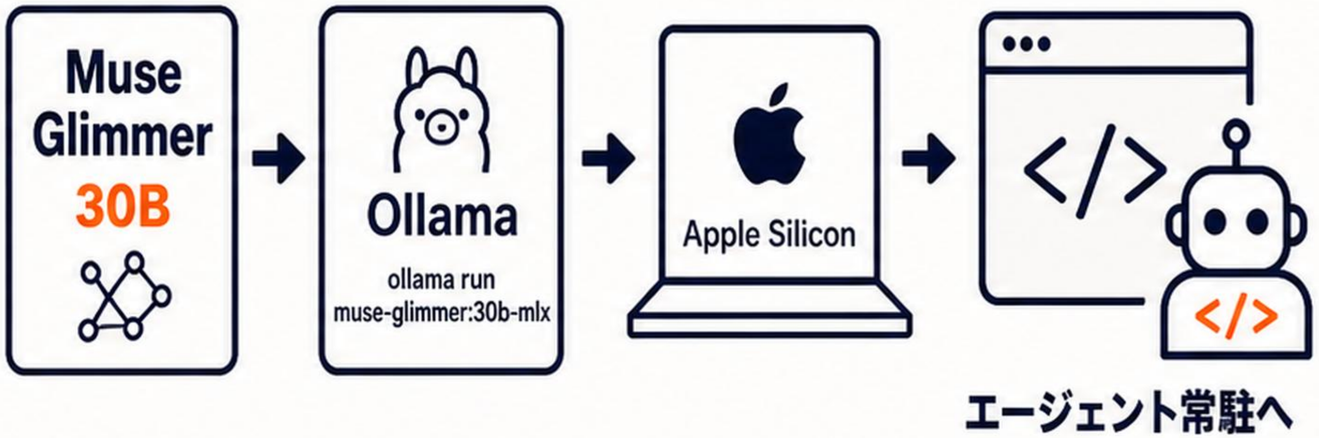
エージェント運用のコストは「賢いモデルで計画し、量を捌くモデルで実装する」構成の後者に集中する。

Meta の新モデル Muse Glimmer が Ollama で動く

30B・Apache 2.0・ローカルのエージェント常駐へ

Meta Superintelligence Labs の初リリースとなる新オープンモデル Muse Glimmer が、Ollama で実行できるようになった。

- **30B** のマルチモーダルモデル、**128K+** コンテキスト、**Apache 2.0**。Meta Superintelligence Labs の初リリース
- **ollama run muse-glimmer:30b-mlx** で起動。**ollama launch** は Claude Code / Codex / OpenCode / GitHub…
- Ollama の MLX エンジンが **DFlash** に対応し、Apple Silicon 上で **1.5〜1.8倍高速化**



ローカルLLMの評価軸が「チャットができるか」から **「エージェントを常駐させられるか」** に移った。

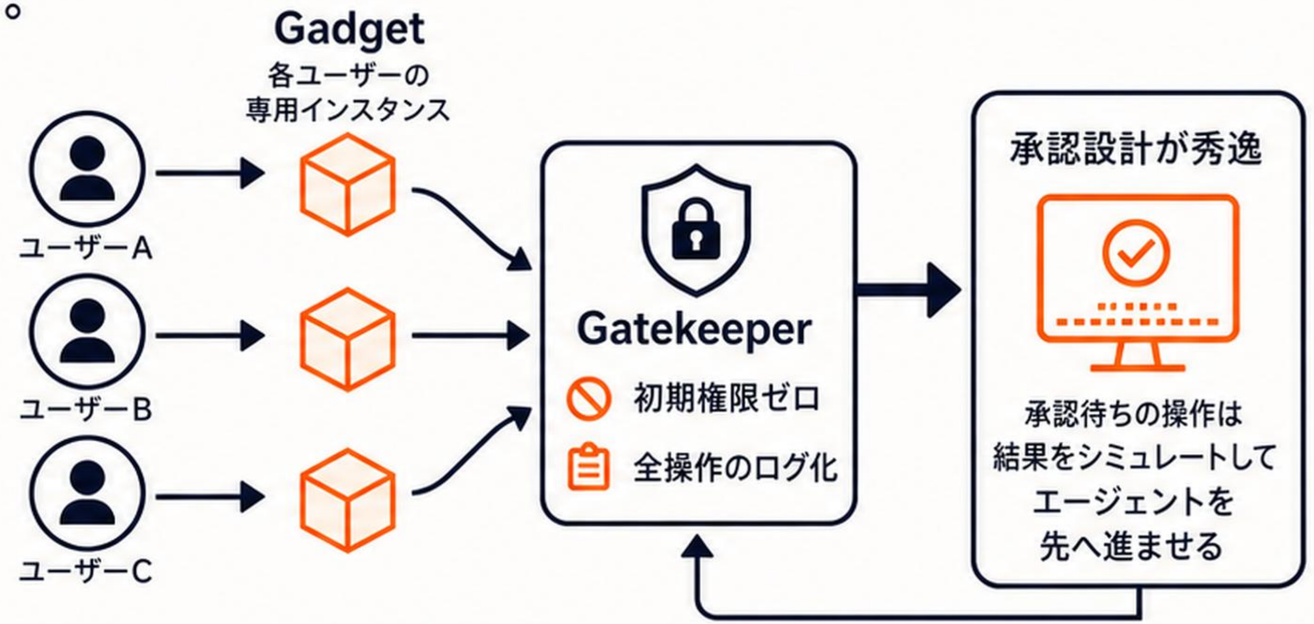
ollama.com

Cloudflare OS をコードから読む

Gadget と Gatekeeper が示す「会社OS」の設計

Cloudflare が社内AI基盤「Cloudflare OS」をオープンソース公開し（8/5）、
それを clone して実際に動かした解説が出てきた。

- **Gadget:** 全ユーザーが**自分専用インスタンス**を持つため、アプリのバグで他人のデータが漏れる構造がそもそ…
- **Gatekeeper: 初期権限ゼロ＋全操作のログ化。** MCP のように全ツールを最初から見せる方式と対比され…
- **承認設計が秀逸:** 承認待ちの操作は結果をシミュレートしてエージェントを先へ進ませ…



議論の軸が「AIをどう賢くするか」から「AIが作ったものをどう統制するか」へ完全に移ったことを示す実装。

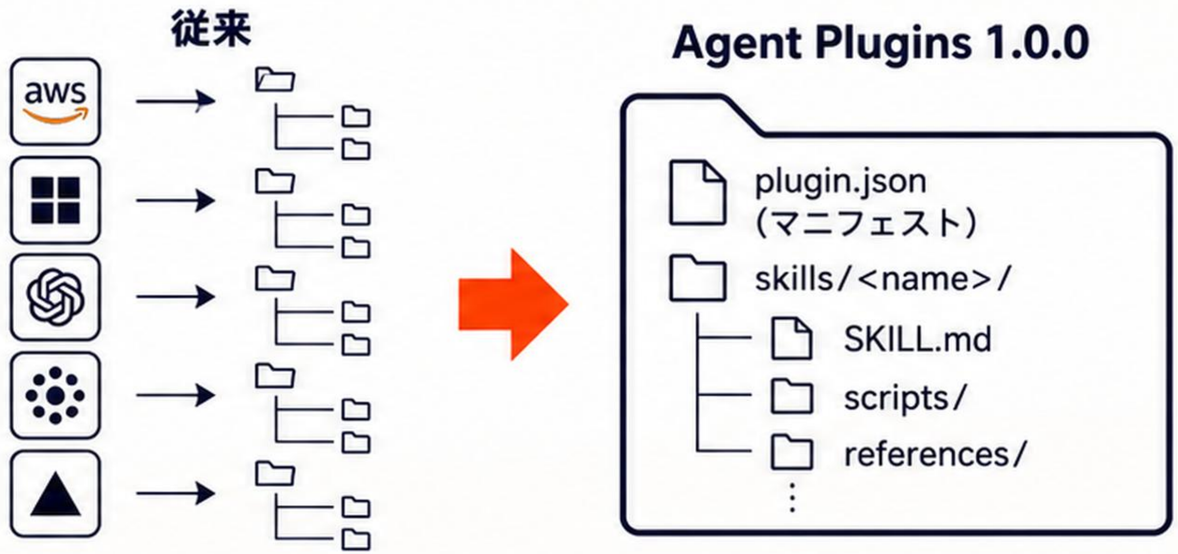
x.com

Agent Plugins 1.0.0 に Google も参加

スキルとMCP設定がベンダ横断で共通化、Anthropic は沈黙

AWS・マイクロソフト・OpenAI・Anysphere・Vercel が策定した AI エージェント向け共通パッケージ仕様「Agent Plugins 1.0.0」に、Google も対応を表明した。

- 策定は AWS・マイクロソフト・OpenAI・Anysphere・Vercel。Google も対応を発表
- フォルダ構成は plugin.json（マニフェスト）/ skills/<name>/{SKILL.md, scripts/, references/} /…
- 従来はベンダごとに設定方法とフォルダ構成が異なり、同じ内容を書き換える必要があった



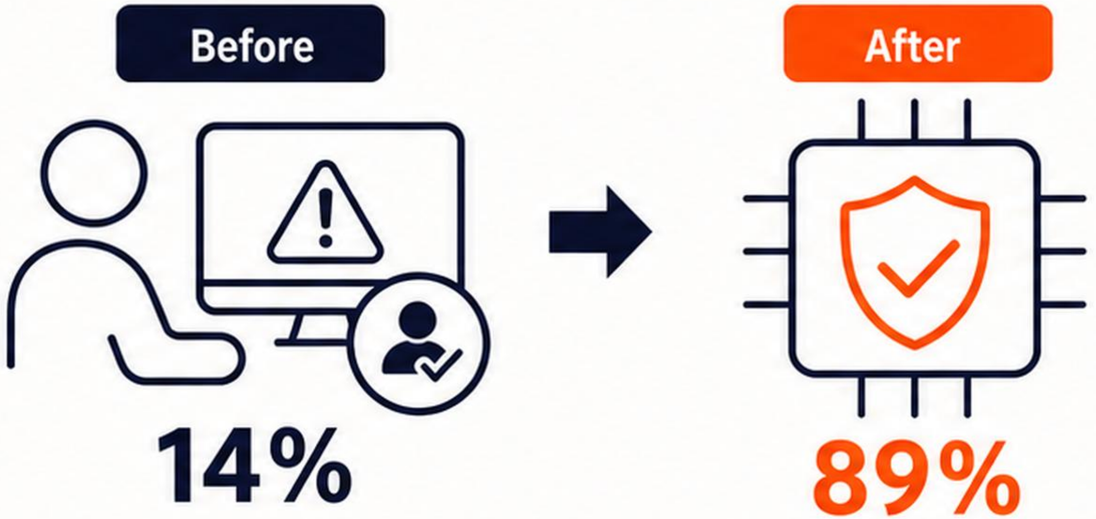
スキルはこれまで「Claude Code のもの」「Cursor のもの」と各ランタイムに閉じていた。

Claude Code の Auto mode、8月14日から既定の権限モードに

分類器が危険コマンドの89%を検出

8月14日から、Claude Code の Pro・Max・Team ユーザーで Auto mode が既定の権限モードになる。

- **危険コマンドの検出率: Auto mode 89% / 手動承認 14%**
- 1,053人の有料テスターを対象に、権限確認プロンプトの文面を明らかに危険なコマンドへ差し替える調査を実施（…
- 変更理由は2つ: ①追跡した全安全指標で Auto mode が手動承認と同等以上…



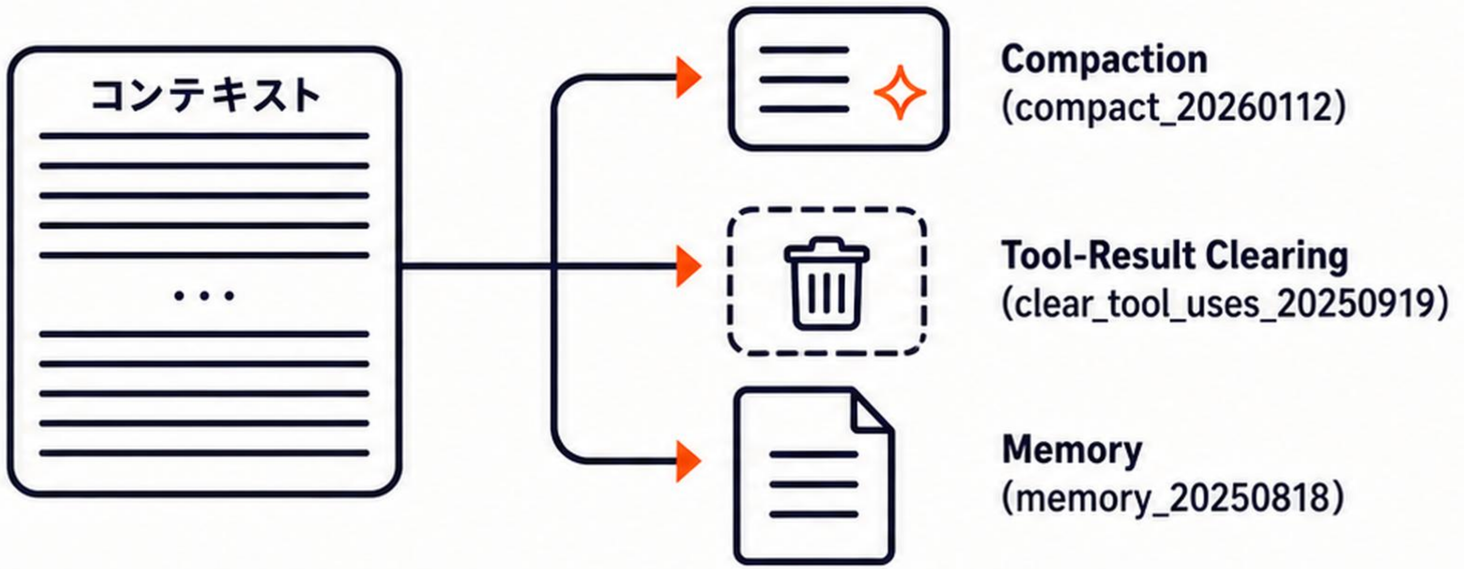
「人間が承認するから安全」という前提を、実測で正面から否定した点が重い。

〈知見〉Anthropic の Context Engineering Cookbook

3プリミティブと実測値

Anthropic が、エージェントのコンテキストを実運用規模で
管理するための手引きを公開した。

- **Compaction (compact_20260112):**
  会話全体を高精度な要約へ圧縮して
  置き換える。推論コストがかかる…
- **Tool-Result Clearing (clear_tool_uses_20250919):**
  tool_result…
- **Memory (memory_20250818):**
  /memories にファイルとして書き出す。
  保存内容は無損失で…



「コンテキストが溢れる」を一括りにせず、何が窓を食っているかで打ち手を変えるという分解が本体。

platform.claude.com

本日のトピック一覧

1. Anthropic、未公開の研究版 Claude がリーマンゼータの下界を 41.6% → 67.2% に更新
2. OpenAI、サイバーセキュリティ専用モデル GPT-5.6-Cyber を投入し Daybreak を拡張
3. Anthropic、Claude Sonnet 5 の導入価格 $2/$10 を恒久化（9/1 からの値上げを撤回）
4. Meta の新モデル Muse Glimmer が Ollama で動く
5. Cloudflare OS をコードから読む
6. Agent Plugins 1.0.0 に Google も参加
7. Claude Code の Auto mode、8月14日から既定の権限モードに
8. 〈知見〉Anthropic の Context Engineering Cookbook

2026-08-11

TEKION Group