Tech News Daily

2026-07-03

MORNING DISPATCH / Vibe Coder Bootcamp Tech News

3 トピックを整理。

TEKION Group

Tech News Daily
2026-07-03

Claude Code v2.1.197〜2.1.198 が大型更新 — Sonnet 5 が既定モデルに（1M）、Claude in Chrome が GA、背景エージェントが自動コミット＆ドラフトPR

OpenAI公式

Likes 393

何が起きた？

Claude Code の 2.1.197〜2.1.198 で、既定モデルが Claude Sonnet 5（ネイティブ1Mトークン、8/31まで $2/$10 の導入価格）に切り替わり、Claude in Chrome が正式版（GA）に到達した。あわせて背景エージェントが自動コミットしてドラフトPRを開く、Explore エージェントが Haiku→Opus に格上げ、組織既定モデル設定、サブエージェントの5階層ネストなど、自動化・運用まわりが一段強化された。

主な変更点

- 既定モデルが Sonnet 5 に。ネイティブ1Mトークン・最大出力128k。adaptive thinking が既定ON、手動 extended thinking は廃止（400エラー）。導入価格は8/31まで $2/$10 per Mtok
- Claude in Chrome が GA。拡張機能連携で Webアプリのテスト・コンソールログでのデバッグ・フォーム入力自動化・ページからのデータ抽出が可能。/chrome →「Select browser…」で使うブラウザを選択
- 背景エージェントが自動コミット＆ドラフトPR作成、短時間のネット断は自動リトライ。agent_needs_input / agent_completed の通知フックでデスクトップ通知が可能に。Explore エージェントは Haiku→Opus へ格上げ
- 組織既定モデル（Org default / Role default）を管理者が設定可能に。サブエージェントがさらにサブエージェント を spawn 可能（最大5階層）
- セキュリティ: claude mcp list/get がリポジトリ自己承認した .mcp.json サーバーを起動しないよう修正、未信頼ワークスペースは「⚠ Pending approval」表示。1Mコンテキストをクレジット無しで使い永続的に詰まる不具合を修正

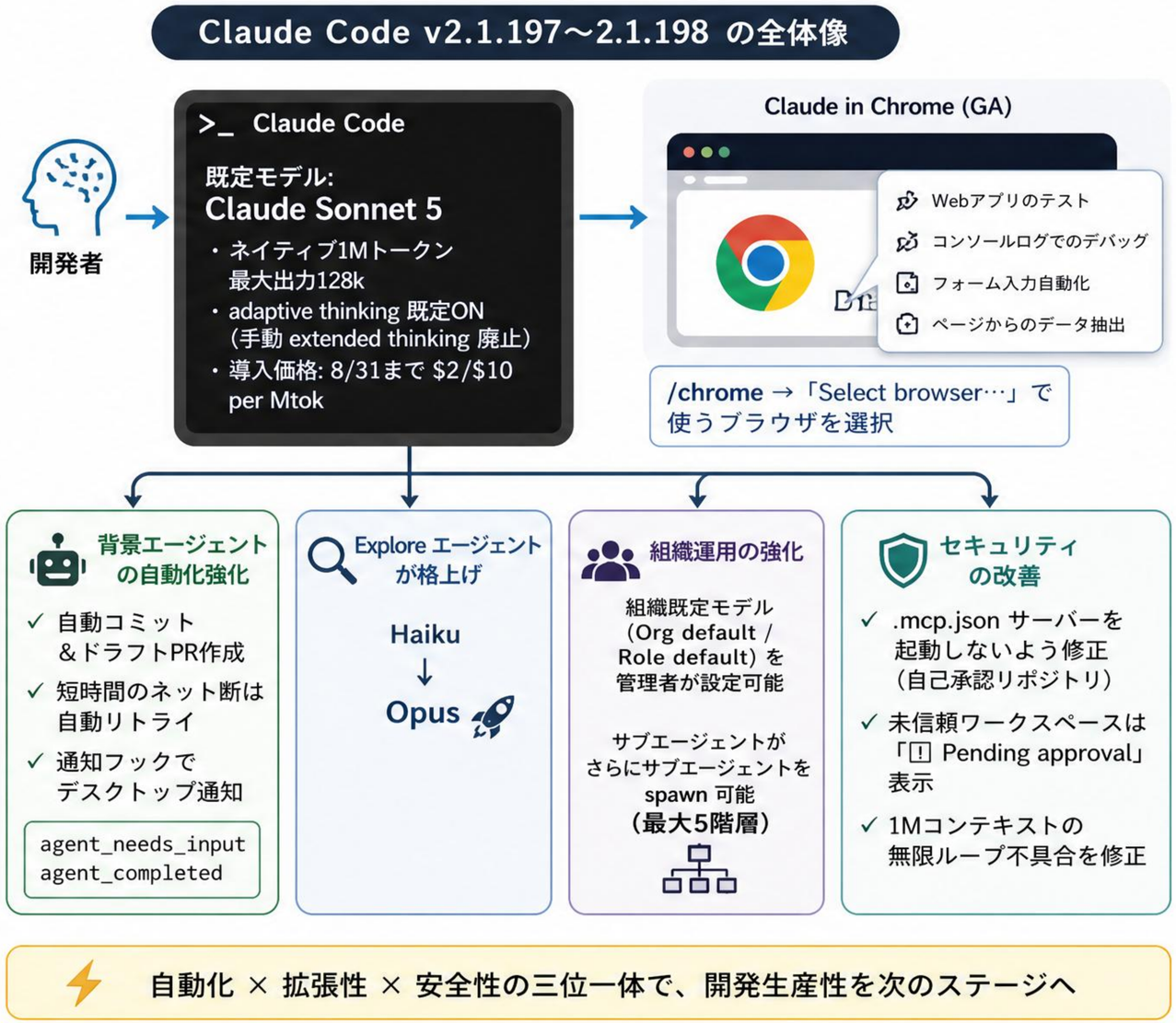
なぜ重要？

Sonnet 5 の 1M コンテキストと標準ONの adaptive thinking により長文・大規模コードでも推論の安定性が向上。Chrome 連携と背景エージェントの自動化でブラウザ操作からコミット・PR までをシームレスに実行できるようになり、開発サイクルを大幅に短縮。組織既定モデルや多階層エージェントにより、チームでの一貫運用とスケーラビリティも強化。セキュリティ修正も含め、AI開発の生産性と信頼性を底上げする大型アップデート。

Xでの反応

@ClaudeCodeLog（changelog bot）の 2.1.197 告知が Likes 393。開発者からは「Chrome 対応と賢い背景エージェント」「unattended run 向けの新コマンド/環境変数」への注目と、既定モデル変更に伴う挙動差への戸惑いの声が混在。

出典: x.com/OpenAI/...



TEKION Group

Agents-A1 公開 — 35B MoE の長時間タスク特化エージェントモデル、パラメータでなく「行動の長さ」をスケールして1兆級に匹敵（Apache-2.0）

OpenAI公式

likes 812

何が起きた？

上海 AI Lab の InternScience グループが、検索・科学研究・ソフトウェア工学・ツール呼び出しなどの long-horizon（長時間・多段）タスクに特化した 35B MoE モデル **Agents-A1** を **Apache-2.0** で公開した。パラメータを増やすのではなく学習軌跡の長さをスケールする「**Horizon Scaling**」により、GPT-5.5 や DeepSeek-V4-pro、Kimi-K2.6 等の frontier 級と互角、一部の長時間ベンチで SOTA を主張している。

主な変更点

- Qwen3.5-35B-A3B ベースの MoE。256Kコンテキスト、1ターン内のツール呼び出し回数に制限を設けない設計。SGLang / vLLM で OpenAI 互換エンドポイント提供
- 「Horizon Scaling」= 平均45Kトークンの長い agentic 軌跡を生成する知識-行動インフラで学習。3段階（全ドメインSFT → ドメイン別教師モデル → マルチ教師のドメインルーテッド on-policy 蒸留）で6ドメインを1つの student に統合
- ベンチ主張（自社・preprint）：GAIA 96.0 / BrowseComp 75.5 / FrontierScience-Olympiad 79.0 / IFEval 94.8 など。35B級では SOTA、frontier 級と競合
- 重みは Hugging Face（Apache-2.0）、コードは GitHub で公開
- ⚠ **preprint 品質でベンチは第三者検証で軟化する可能性あり。**自ワークロードでの独立評価を推奨

なぜ重要？

これまでのスケール競争は「パラメータ数」を増やすことが主流だったが、Agents-A1 は「行動の長さ（学習軌跡）」をスケールする新しいアプローチで、長時間・多段のエージェントタスクにおける性能を大きく引き上げた点が革新的。35B クラスのモデルで frontier 級に匹敵する性能を示したことは、コスト効率の高い エージェント時代の到来を示唆する。オープンなライセンス（Apache-2.0）とエコシステム（HF / GitHub 公開）も魅力。

Xでの反応

「35B で trillion 級に匹敵」のインパクトに注目が集まる一方、「Qwen3.5-35B の post-trained 版として要検証」「preprint なのでベンチは割り引いて見るべき」という慎重な声も。

Agents-A1 の全体像

Agents-A1
35B MoE モデル
long-horizon（長時間・多段）タスクに特化

検索 / 科学研究 / ソフトウェア工学 / ツール呼び出し / …

Qwen3.5-35B-A3B ベースの MoE | 256K コンテキスト | ツール呼び出し回数：制限なし | SGLang / vLLM（OpenAI 互換エンドポイント）

Horizon Scaling の学習フロー

平均45Kトークンの長い agentic 軌跡を生成する知識-行動インフラで学習

1. **全ドメインSFT** — 6ドメインの多様なデータで基礎能力を獲得
2. **ドメイン別教師モデル** — 各ドメインに特化した教師モデルを構築
3. **マルチ教師のドメインルーテッド on-policy 蒸留** — マルチ教師の知識を統合し1つの student に蒸留

統合対象の 6ドメイン
- 検索
- 科学研究
- ソフトウェア工学
- ツール呼び出し
- データ分析
- 一般知識・対話

主なベンチ主張（自社・preprint）

GAIA	BrowseComp	FrontierScience-Olympiad	IFEval	35B級では SOTA、frontier 級と競合
96.0	75.5	79.0	94.8	

重み：Hugging Face（Apache-2.0）
コード：GitHub で公開

⚠ preprint 品質でベンチは第三者検証で軟化する可能性あり。自ワークロードでの独立評価を推奨

Tech News Daily
2026-07-03

Gemini 3.5 Flash に「コンピュータ操作（computer use）」が標準搭載 — ブラウザ/モバイル/デスクトップを1モデルで操作、専用モデル不要に

OpenAI公式
likes 101

何が起きた？

Google が Gemini 3.5 Flash に computer use（画面を見て操作する）を標準ツールとして統合した（public preview）。これまで別モデル（2025-10 の Gemini 2.5 computer use）だった機能が、関数呼び出し・Search grounding・Maps と同じ本番モデルにネイティブ搭載され、ブラウザ・モバイル・デスクトップを1回のモデル呼び出しから操作できるようになった。

主な変更点

- スクリーンショット取得 → Flash がピクセルを読み次手を計画 → 正確なUIコマンド（クリック座標・入力）を出力 → 実行して再スクショ、のループで動作。専用モデルへのルーティングが不要に
- 対応3環境: Webブラウザ / モバイル（タッチ入力シミュレート）/ デスクトップソフト
- ベンチ OSWorld-Verified で **78.4**（16モデル中5位。GPT-5.5 は78.7、Anthropic Opus 4.8 が83.4 で首位）。⚠スコアは各社自己申告・第三者未検証
- 安全策: prompt injection 対策の敵対的学習に加え、機微/不可逆操作にユーザー確認を必須化 / 間接プロンプトインジェクション検知で自動停止、の2つの企業向けセーフガードを任意（opt-in）で提供
- 価格: computer use 利用に追加課金なし。Gemini API / Gemini Enterprise Agent Platform で提供、Browserbase デモと GitHub 参照実装あり

なぜ重要？

これまで専用モデルへの切替や別基盤の用意が必要だった「画面を見て操作する」機能が、本番モデルの標準機能として統合。ブラウザ・モバイル・デスクトップを1つのモデルで横断操作できるため、開発・運用がシンプルになり、エージェント体験の一貫性と拡張性が大きく向上する。

Xでの反応

「Gemini 3.5 Flash に手が生えた（computer use 標準搭載）」として拡散。専用モデル切替が不要になった実装面の楽さを評価する声が中心。

RoundtableSpace
Likes 101

Gemini 3.5 Flash の computer use 仕組み（ループ）

対応3環境を1モデルで操作

安全策（opt-in）

ユーザー確認を必須化
機微/不可逆操作の前にユーザー確認を挟む

間接プロンプトインジェクション検知で自動停止
不正な指示の影響を検知し、自動で停止して安全を確保

ベンチマーク（OSWorld-Verified）

78.4（16モデル中5位）

GPT-5.5: 78.7
Anthropic Opus 4.8: 83.4（首位）

⚠ スコアは各社自己申告・第三者未検証

価格

追加課金なし
Gemini API / Gemini Enterprise Agent Platform

提供リソース

Browserbase デモと GitHub 参照実装あり

出典: x.com/OpenAI/...

TEKION Group

本日のトピック一覧

1

Claude Code v2.1.197〜2.1.198 が大型更新 ―
Sonnet 5 が既定モデルに（1M）、Claude in Chrome が GA、背景エージェントが自動コミット＆ドラフトPR

claude-code を v2.1.197 に更新。Sonnet 5 が既定モデルに（1M コンテキスト）。
Claude in Chrome が GA。背景エージェントが自動でコミット＆ドラフト PR を作成。

2

Agents-A1 公開 ― 35B MoE の長時間タスク特化エージェントモデル、パラメータでなく「行動の長さ」をスケールして1兆級に匹敵（Apache-2.0）

35B MoE ベースのエージェントモデル「Agents-A1」を公開。パラメータ数ではなく、
「行動の長さ」をスケールすることで 1兆パラメータ級モデルに匹敵する性能を実現（Apache-2.0）。

3

Gemini 3.5 Flash に「コンピュータ操作（computer use）」が標準搭載 ―
ブラウザ/モバイル/デスクトップを1モデルで操作、専用モデル不要に

Gemini 3.5 Flash に「コンピュータ操作（computer use）」が標準搭載。
ブラウザ、モバイル、デスクトップの操作を 1 つのモデルで実現し、専用モデルが不要に。

出典一覧

OpenAI公式

ANTHROP\C
Anthropic公式

Google公式

TEKION Group

作成日：2025年8月25日
TEKION Group Inc.