

TRAINING

生成AI 最新トレンド

2026年7月時点

2026年7月28日 時点の公開情報にもとづきます

講師 安田 光喜 (Givery)



本資料の二次転載禁止について

本資料は株式会社ギブリーが作成した研修教材で、著作権はギブリーに帰属します。

お願い

社内での受講と復習を目的とした閲覧・印刷は問題ありません。社外への配布、SNSやブログへの掲載、他社研修への流用はご遠慮ください。引用が必要な場合は担当者までご連絡ください。

統計と方針は公開情報にもとづき、出典は各ページに記載しています。

この資料の見方

この資料の数字は**2026年7月28日**までの公表値だけです。

◆ **すべて2026年7月時点の断面です**

モデルの順位もベンチマークの構成も数か月で入れ替わります。

第4章の総合指標の表は継続更新されるので、投影日の数字は違う前提で見てください。

◆ **数値には基準日と出典を添えています**

同じ指標でも測った日と足場が違えば値が変わります。SWE-bench Verifiedは公式ハーネスで76.80%（2026年2月17日）ですが、ベンダー自己申告では90%台が出回っています。

◆ **報道が食い違う数値は断定しません**

Claude Design公開当日のFigma株の下落率は、媒体により6%から7.7%まで幅があります。

この資料では幅のまま出すか、載せません。

幅で書く理由

食い違う数字を1つに丸めると、後で反証されたときに資料全体が疑われます。

全体の流れ

7章のうち、選定の判断に直結するのは**第4章と第7章**です。

章	テーマ	持ち帰るもの
第1章	現在地	単価で選ぶ段階に入ったという事実
第2章	インパクトニュース	Kimi K3とClaude Designが変えた前提
第3章	ゲームとエンタメ	AI開示が売上に効くという数字
第4章	モデル性能比較	用途別の使い分けの基準
第5章	ベンチマークの世代交代	自社で評価を作る必要が出た理由
第6章	国内企業の活用事例	公表値の読み方と過大評価の回避
第7章	勢力図と国家	計算資源と規制が選定に効く経路



01

2026年7月の現在地

総合指標の上位3モデルは2ポイント差に収まりました。差がつくのは1タスクあたりのコストと、重みが公開されているかどうかです。

性能の差ではなく、1タスクいくらかで モデルを選ぶ段階に入りました。

総合指標の上位はClaude Opus 5が61、Claude Fable 5が60、
GPT-5.6 Solが59です。上位3モデルの差は2ポイントです。
同じ表の1タスク平均コストは最安\$0.04から
最高\$2.75まで開き、その差は約69倍です。

2026年前半の主要発表

この7か月で実務に効いたのは、**重みの公開日と規制の適用日**です。

日付	出来事	何が変わるか
1月16日	ValveがSteamのAI開示ルールを改定	開示義務の対象を絞る
2月19日	パナソニック コネクトがAIエージェント運用開始	図面と設計仕様の照合
4月17日	AnthropicがClaude Designを公開	同日にFigma株が6～7.7%下落
6月9日	Claude Fable 5とMythos 5が提供開始	総合指標の首位が交代
7月9日	AtCoder世界一決定戦でAIが人間に勝利	去年は人間が優勝していた
7月16日	Moonshot AIがKimi K3を発表	7月27日に重みを公開
7月22日	ホワイトハウスがGenesis Mission発表	連邦15機関超で50億ドル超
7月24日	AnthropicがClaude Opus 5を公開	総合指標61で首位

この先の確定日程は2026年8月2日、EU AI Act第50条の透明性義務が適用開始です。

現在地を示す3つの数字

同じ表の中で、性能差は2ポイント、コスト差は69倍です。

2 ポイント

総合指標の上位3モデルの差
(2026年7月28日時点)

4 か月

オープンウェイトがクローズドに
遅れる期間 (2026年5月28日時点)

69 倍

1タスク平均コストの上下の開き
(2026年7月28日時点)

内訳

上位3モデルはClaude Opus 5が61、Claude Fable 5が60、GPT-5.6 Sol (max) が59です。1タスク平均コストはDeepSeek V4 Proの\$0.04からClaude Fable 5の\$2.75まで開きます。オープンウェイトとの差はEpoch AIの計測で平均4か月、ECIで8ポイントです。

出典 artificialanalysis.ai/models、epoch.ai/data-insights

02

インパクトニュース

競技プログラミングでの逆転、Kimi K3、デザイン領域への進出。7月に動いた3件を数字で押さえます

昨年は人間が勝った大会で、**1年で逆転**しました。

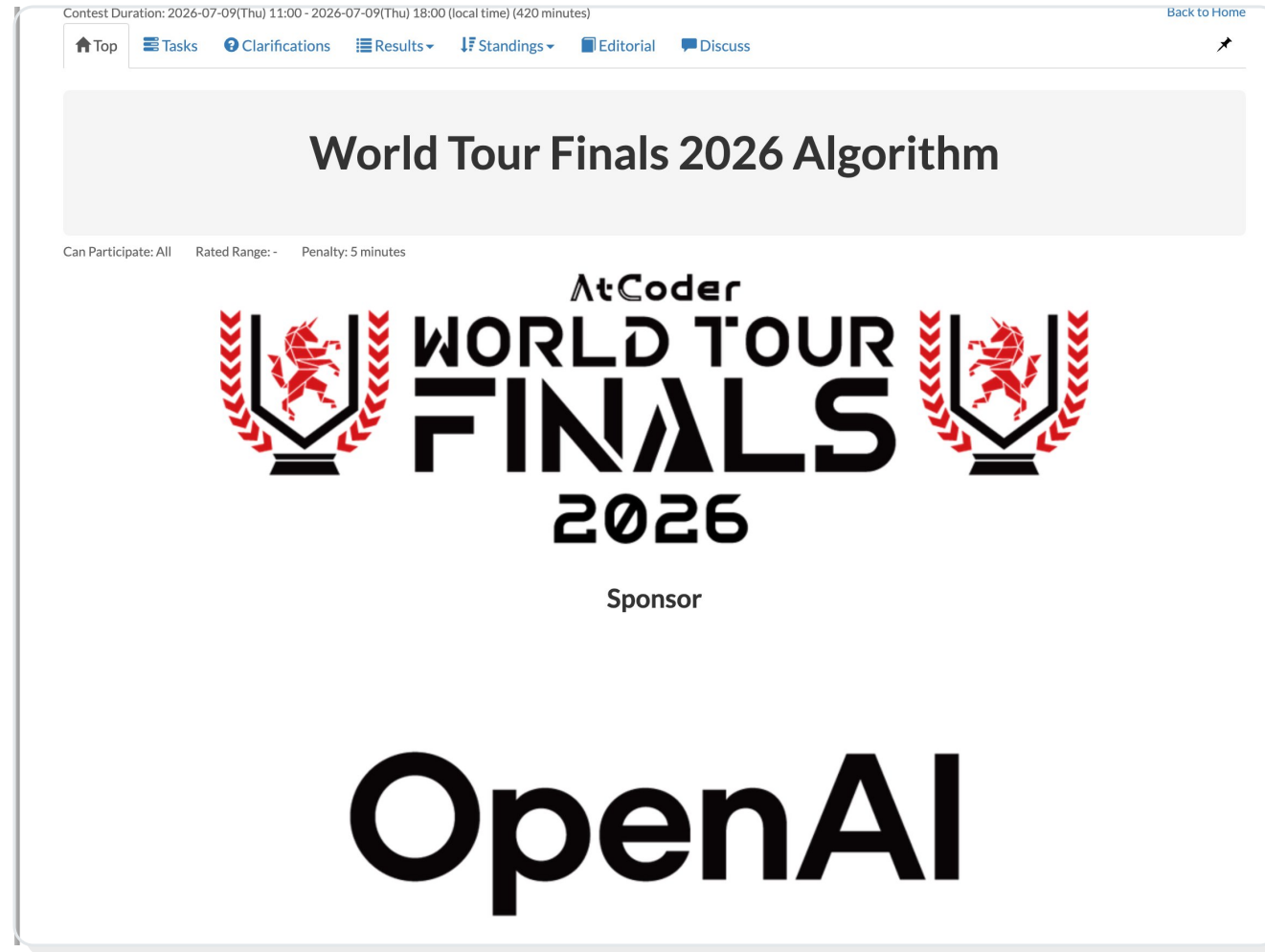
項目	内容
大会	AtCoder World Tour Finals 2026（Heuristic 7月7～8日／Algorithm 7月9日）
AIの参加形態	OpenAIがエキシビジョンとして参加。出場モデルの正式名は非公表
人間の出場者	各部門12人。1年かけて予選を勝ち上がった世界の上位
Algorithm部門	AIがA～Eの5問すべてを解答。制限時間は420分
人間トップ	tour1st がD問題まで。最後のE問題で及ばず
Heuristic部門	序盤からAIが引き離して勝利。2025年はOpenAIが2位

効いているのは勝ったことではなく、勝ち方です

発想の質で人間を上回った点が想定を超えたと報じられています。スコア差は非公表です。

出典 chokudai.hatenablog.com、understandingai.org、日経クロステック（いずれも2026年7月）

AtCoder World Tour Finals 2026



公式ページの実画面。開催は2026年7月9日の420分、スポンサーはOpenAI（atcoder.jp、2026年7月29日取得）

Kimi K3 の構成

重みが公開されたモデルが、商用最上位と4ポイント差に来ました。

項目	内容
発表と重み公開	2026年7月16日に発表、7月27日に Hugging Face で重みを公開
パラメータ	総2.8兆、アクティブ1,040億（896専門家中16を選択）
コンテキスト長	1,048,576トークン。前世代 K2.6 の256Kから4倍
入出力	テキストに加えて画像と動画をネイティブに扱います
知能指数	57。首位 Claude Opus 5 は61（7月28日）。公開当時の7月17日は全体3位

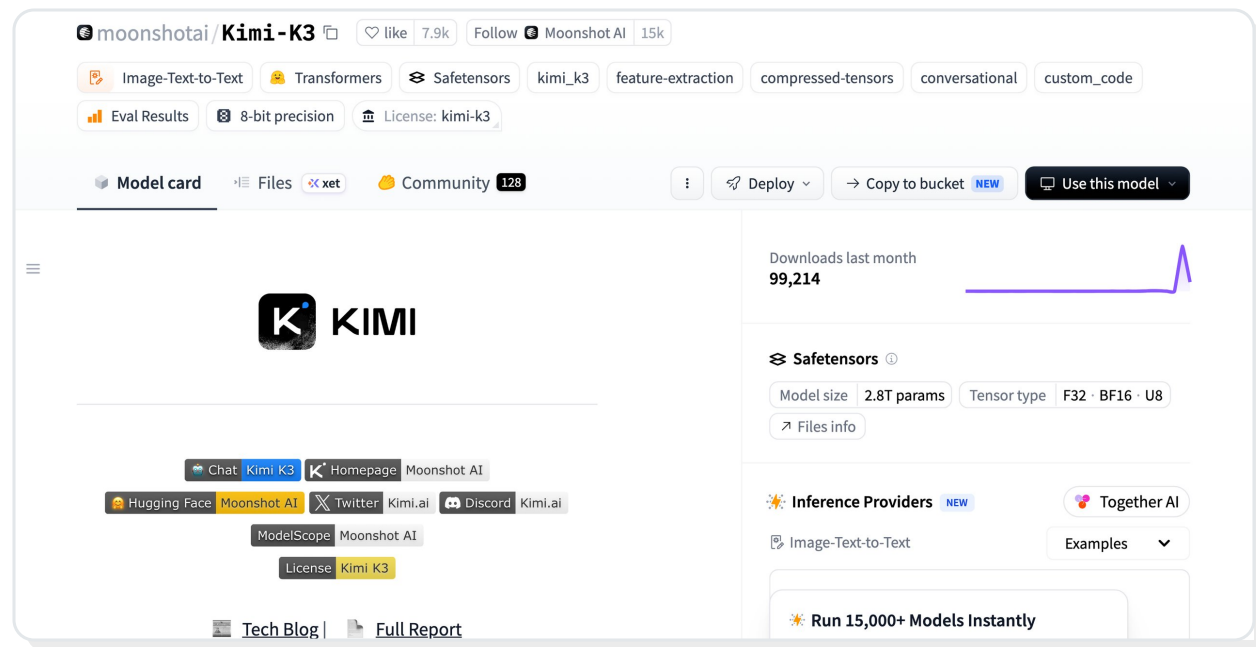
なぜニュースなのか

ダウンロードできる重みが商用最上位に3ポイントまで迫りました。性能を落とさずに自社ホストを選べる状態は、2026年7月が最初です。アナリストは Fable 級の中国モデルの登場を2027年初頭と見込んでいました。

出典 kimi.com/blog/kimi-k3、huggingface.co/moonshotai/Kimi-K3、artificialanalysis.ai

重みの公開とライセンス

誰でも落とせますが、誰でも売れるわけではありません。



moonshotai/Kimi-K3 モデルカード（Hugging Face、7月27日時点）

ライセンス

MIT を改変した独自ライセンスです。直近12か月の売上が2,000万ドルを超える事業者がモデルをサービス提供する場合は、Moonshot AI との個別契約が必要です。月間1億ユーザーを超える製品は UI への「Kimi K3」表示が求められます。

公開直後の反応

直近1か月のダウンロードは99,214件、いいねは7.89千件です（2026年7月27日時点）。

性能の位置

上位4モデルの知能指数は57から61の間です。



Artificial Analysis 知能指数（2026年7月17日算定）。目盛りは0から60。

長時間の仕事で測ると

非公開評価 AA-Briefcase（2026年7月21日）の Elo は Kimi K3 が1543、Claude Fable 5 が1574、GPT-5.6 Sol（max）が1501。前世代 Kimi K2.6 は816で、1世代で727ポイント上がりました。ただし1タスクの所要は56.4分で Fable 5 の約2.5倍、タスク単価は10.57ドルかかります。

出典 artificialanalysis.ai（知能指数・AA-Briefcase）

価格インパクト

他社比では3分の1、自社比では約4倍の値上げです。

100万トークンあたり	Kimi K3	Claude Fable 5
ブレンド単価（AA算定 7:2:1）	\$2.31	\$7.70
出力単価	\$15.00	\$50.00

反対側の事実

前世代 Kimi K2.6 の単価は入力\$0.95・出力\$4.00でした。K3 は入力\$3.00・出力\$15.00で、自社の前世代からは約4倍です。安い中国製モデルという前提は、ここで崩れました。

出典 platform.kimi.ai、artificialanalysis.ai（2026年7月27日時点）。
Fable 5 の出力\$50は Fortune の報道値です。

Flagship Model Kimi K3 Pricing

Copy page

Product Pricing

Explanation: Prices exclude applicable taxes. Specific tax obligations are subject to local tax regulations and will be calculated at checkout based on your jurisdiction.

Model	Unit	Input Price (Cache Hit)	Input Price (Cache Miss)	Output Price	Context Window
kimi-k3	1M tokens	\$0.30	\$3.00	\$15.00	1,048,576 tokens

Here, 1M = 1,000,000. The prices in the table represent the cost per 1M tokens consumed.

Kimi API Platform の料金表

使う前の落とし穴

置き換えの話ではなく、**どこに置くか**の話です。

◆ 重みは1.56TB・96シャード、単一GPUには載りません

vLLM の推奨構成は 8×B300 の1ノード、または 16×B200 です。試すだけでも計算基盤の調達から始まります（2026年7月27日時点）。

◆ 同じ\$3/\$15でも出力速度は5倍違います

Moonshot 純正 API は32.4トークン/秒、Fireworks は163.6トークン/秒。9社が同一単価で提供しており、選ぶ基準は速度になります。

◆ ライセンスに商用条項があります

売上2,000万ドル超の事業者は個別契約、月間1億ユーザー超の製品は表示義務。契約審査を通してから検証に入ってください。

実際の置き場所

Fortune の報道では、DoorDash が難易度の低い処理を Kimi K2.6 に振り分け、Cursor が自社モデル Composer 2 の土台に Kimi を使っています（2026年7月16日）。前段の仕分けが、いまの置き場所です。

出典 vllm.ai/blog/2026-07-27-k3、artificialanalysis.ai、fortune.com

Claude Design で作れるもの

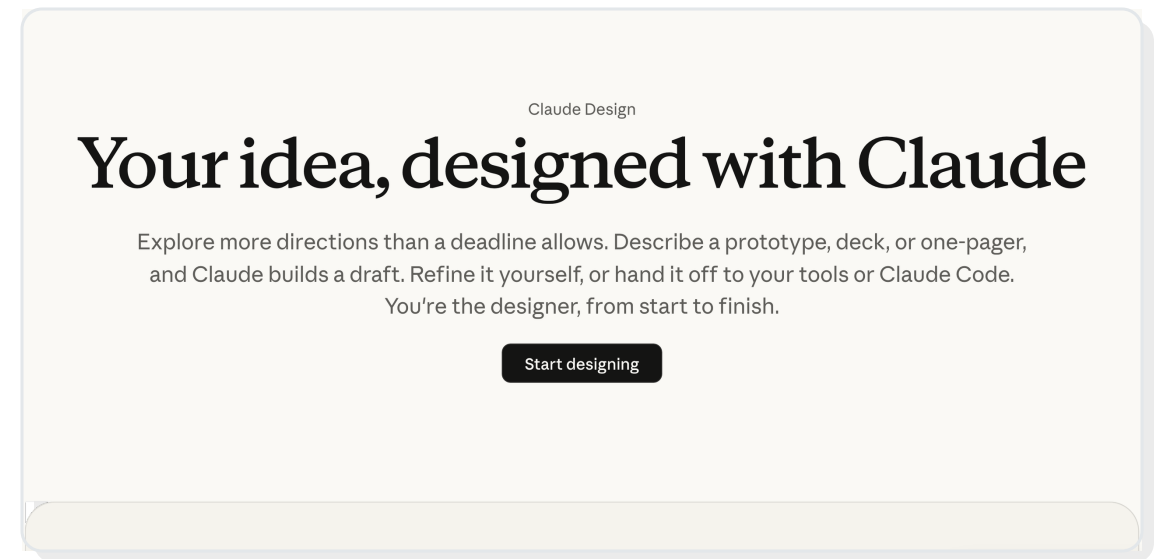
デザインの**前工程**が Claude の中で完結します。

公開時点

2026年4月17日、Anthropic Labs の最初の公開プロダクトとして research preview で公開。Claude Pro / Max / Team / Enterprise 向けで、Enterprise は管理者が有効化します。

作れるもの

- ◆ プロトタイプ
- ◆ ワイヤフレーム
- ◆ ピッチデッキ
- ◆ 1ページ資料



claude.com/product/design (2026年7月時点)

利用の広がり と 連携先

初週の利用者は**100万人**を超えました。

100 万人

公開初週の利用者数

Anthropic の自己申告値で、第三者の検証はありません（2026年6月17日公表）。

区分	内容
コネクタ	Adobe、Base44、Canva、Gamma、Lovable、Miro、Replit、Vercel、Wix
書き出し	PDF、PowerPoint、HTML、組織内リンク

コネクタ一覧に Figma はありません（claude.com/product/design、2026年7月時点）。

業界に出た数字

株価は下げ、売上は46%伸びました。

株価の反応

公開当日の2026年4月17日に約7%下落しました。下落率は報道により6%～7.7%と幅があります。2026年7月28日の株価は24.38ドルで、52週高値142.92ドルから約83%下です。

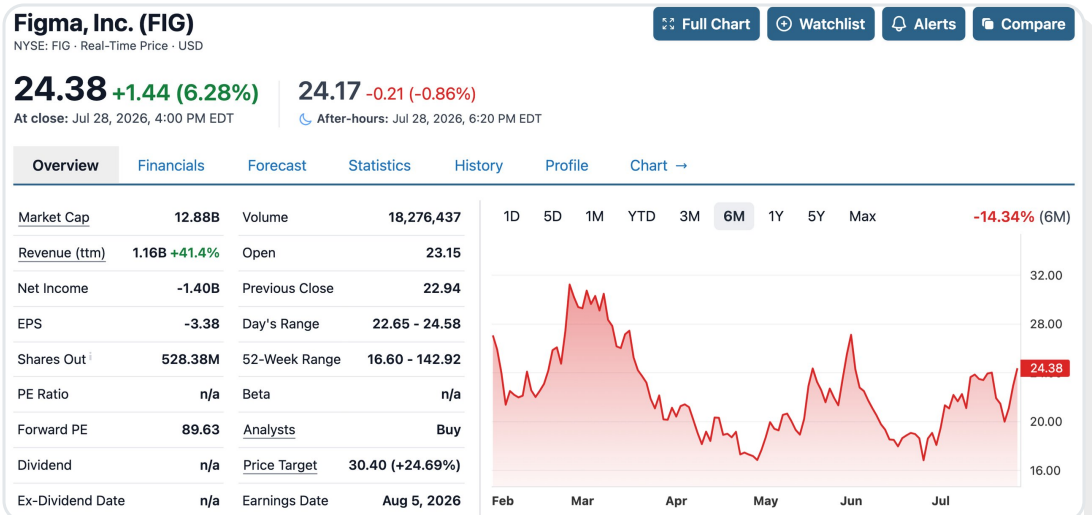
実際の業績

2026年5月15日発表のQ1売上は3億3,340万ドルで前年同期比46%増。通期ガイダンスを5,500万ドル引き上げ、ネットダラーリテンションは139%です。

読み方

通年の下落は2026年を通した AI 競合懸念によるもので、Claude Design 単独の影響ではありません。決算に置き換えの跡は出ていません。

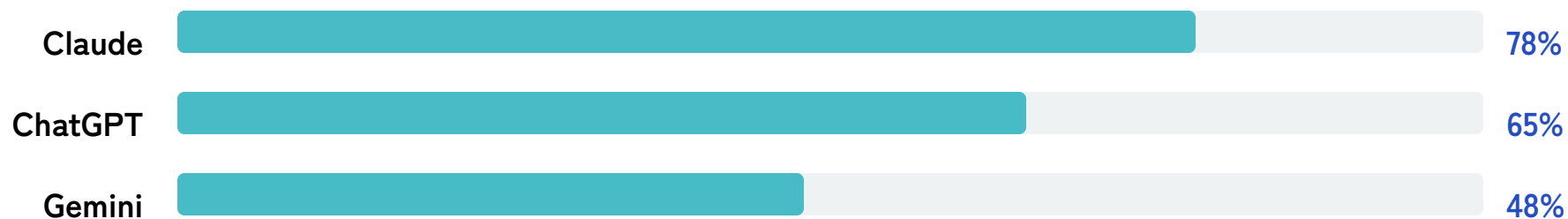
FIG 株価チャート（stockanalysis.com）
株価は2026年7月28日時点。出典 finance.yahoo.com



図中の Revenue (ttm) +41.4% は直近12か月ベース。本文の46%増はQ1単四半期の前年同期比です。

デザイナーのAIツール利用率

デザイナーの**78%**が Claude を使っています。



回答者に占める利用率。目盛りは0から100%。

この数字の読み方

調査は2026年Q1に実施され、公開は7月23日です。Claude Design は調査の締切後に出たので、この78%に Claude Design 自体の利用は入っていません。AIを週1回以上使うデザイナーは91%で、2025年の54%から伸びました。

出典 stateofaidesign.com/chapters/tools (AI in Design Report 2026)

置き換わったのはツールではなく、 デザインの**前工程**です。

Figma の売上は Claude Design 公開後の四半期も46%増、ネットダラーリテンションは139%でした。Anthropic のコネクター一覧に Figma はなく、Figma 側は2026年1月26日に Claude 連携を出しています。効いているのは初期案づくりです。

Anthropic が挙げた顧客証言は「他ツールで20プロンプト以上かかったページが2プロンプト」（Brilliant）でした。ただ、書き出しと複数人同時編集がなく、variants や auto-layout を Figma 並みに扱えないという実測レビューが2026年5月18日に出ています。仕上げの工程は、まだ動いていません。

出典 finance.yahoo.com、anthropic.com/news/claude-design-anthropic-labs、yiannisantoniou.substack.com

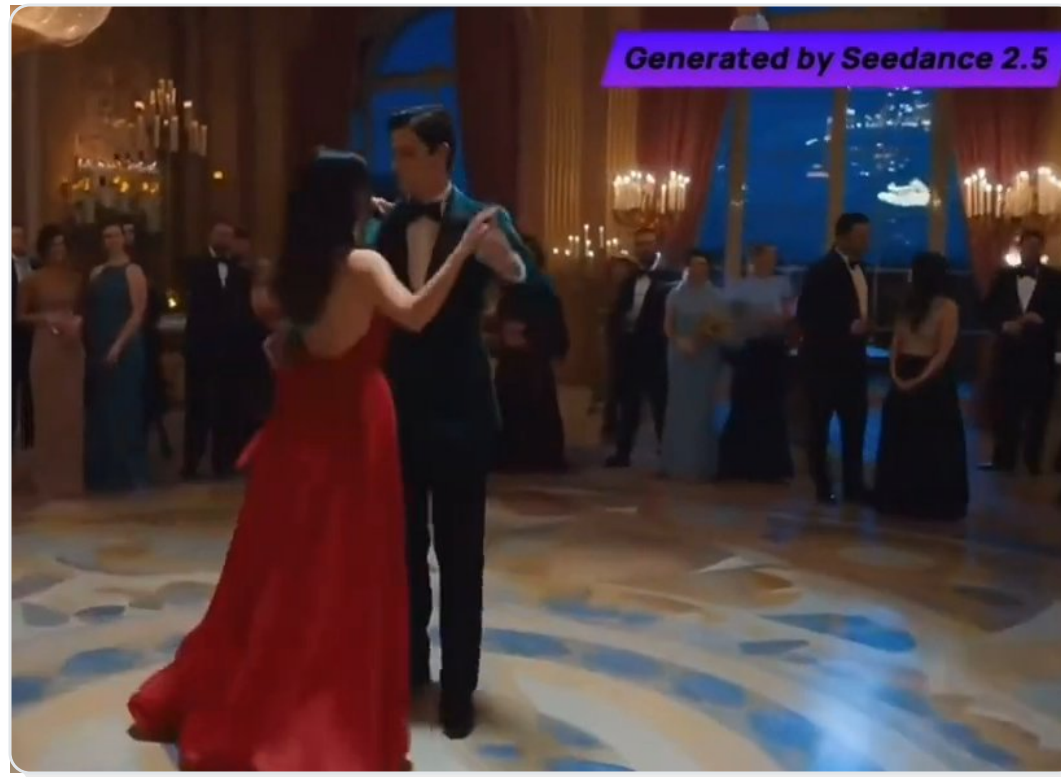
03

ゲームとエンタメ

制作現場では常用段階に入り、同時に開示のコストが数字で見え始めました

動画生成の到達点

群衆も衣装も、実写と見分けがつかない水準にきました。

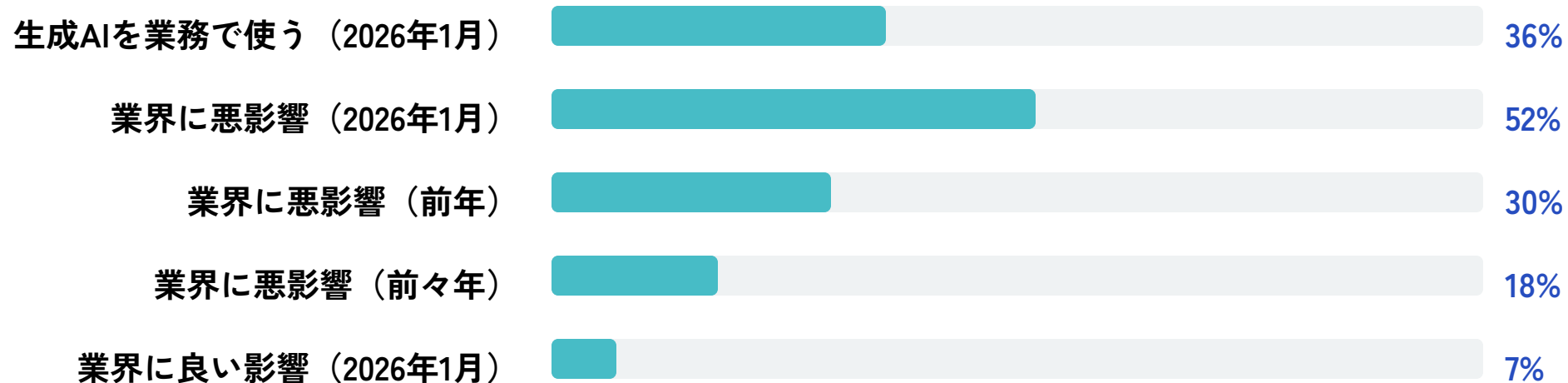


Seedance 2.5 公式デモ（透かしあり）

当日は [videos/seedance-cinematic.mp4](#) を再生します。画面右上の透かしで出力元のモデルが分かります。

制作現場の温度

使う人も、悪影響だと答える人も、**同じ年に増えました。**



出所: GDC 2026 State of the Game Industry (gdconf.com、回答者2,300名超)

所属で割れています

ゲームスタジオ勤務は30%、パブリッシャーや支援・マーケティング系は58%が業務で使っています。二次媒体には利用率を29%とする記事もあり、増減の断定は避けます。

AI開示のコスト

作るのは楽になりましたが、開示すると届きにくくなりました。

指標	数値	時点と調査対象
2026年新作のAI開示率	約3本に1本（約33%）	2026年7月・Steam 53,597本の全数調査
AI開示ゲームのレビュー数	52.6%減（47.69～57.63%）	2025年12月・9,879本の回帰分析
AI開示ゲームの売上シェア	10～27%	2026年7月・同じ全数調査
Steam Next Fest 出展作の開示率	19.71%（1,715本）	2026年6月・約8,700本中・SteamDB

非対称の中身

リリース数のシェアは3割強なのに、売上シェアは1～3割の下限に寄り、レビュー数は統計統制後で半分以下です。開示率は二次媒体が30.8%、一次資料は「約3本に1本」と表記に幅があります。

出所: fragwyz.substack.com / game-oracle.com / pcguide.com

音楽の供給と消費

供給は倍々で増え、消費はほとんど動いていません。

アップロード（供給）

- ・ 1日9万曲が完全AI生成
- ・ 新規アップロードの5割超
- ・ 2026年1月は6万曲・39%、4月は7.5万曲・44%

再生（消費）

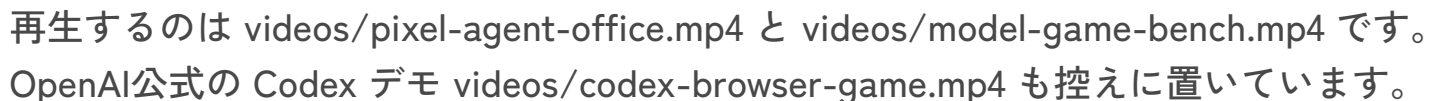
- ・ 再生数に占めるAI楽曲は1～3%
- ・ AI楽曲の再生の最大85%が不正再生
- ・ 不正再生分は収益化されない

1～3%

再生数に占めるAI楽曲の比率（Deezer、2026年7月21日）

アップロードの5割超・1日9万曲が完全AI生成でも、聴かれる量はこの水準です。

投影は静止画のまま、再生はローカルの動画ファイルで行います。



04

モデル性能比較

2026年7月時点の主要モデルを、性能と1タスク当たりコストの両面で並べます

総合指標の上位

上位3モデルの差は2ポイントです。

モデル	知能指数	出力単価	コンテキスト長
Claude Opus 5 (max)	61	\$25	1M
Claude Fable 5	60	\$50	1M
GPT-5.6 Sol (max)	59	\$30	1,050,000
Kimi K3	57	\$15	1,048,576
Grok 4.5 (high)	54	\$6	500k

オープンウェイトのKimi K3が57で4位に入り、1Mトークンは最上位帯の標準になりました。
コンテキスト長は各社の公表表記のままです。1M は100万トークン級の丸め、GPT-5.6 と Kimi K3 は実数で公表しています。
出典 Artificial Analysis Intelligence Index v4.1（2026年7月28日時点）／出力単価は100万トークンあたりの各社公式定価

コーディング性能の比較

実務寄りの評価はAnthropic勢が抜けています。

SWE-Bench Pro（実務規模の改修を任せたときの解決率）



Terminal-Bench 2.1 では順位が入れ替わります

GPT-5.6 Sol Ultra 91.9%、Sol 88.8%、Claude Mythos 5 88% の順です。同じベンチマークをxAIは Fable max 84.3%、Grok 4.5 83.3% と報告しており、走らせる足場が違えば1~2ポイントは動きます。

出典 OpenAI 「GPT-5.6」 発表ページ（2026年7月9日）。各社が自社ハーネスで走らせた値です。

1タスク単価と出力トークン効率

上位2ポイント差なら、**決め手はコスト**です。

モデル	知能指数	1タスク当たり平均コスト
Claude Fable 5	60	\$2.75
Claude Opus 5 (max)	61	\$2.03
GPT-5.6 Sol (max)	59	\$1.54
Kimi K3	57	\$0.72
DeepSeek V4 Pro (max)	44	\$0.04

指数の差は17ポイント、単価の差は約69倍です。

消費トークンでも約4.2倍の開き

平均出力トークンは Grok 4.5 が15,954、Claude Opus 4.8 (max) が67,020（xAI自社計測）。

出典 Artificial Analysis（2026年7月28日時点）／xAI「Grok 4.5」（2026年7月16日、自社計測）

用途ごとの選び方

用途ごとに見る指標が違います。

長時間の調査・分析
途中で止まらない持続力



知能指数が最上位の帯
Claude Opus 5 61、Claude Fable 5 60

実務のコーディング
既存コードへの改修が中心



SWE-Bench Pro の上位
Claude Mythos 5 80.3%、Fable 5 80%

大量処理でコストが効く
単価が数十倍変わる領域



1タスク\$0.5以下の帯
DeepSeek V4 Pro \$0.04、Kimi K3 \$0.72

自社環境に置く
重みを手元に置きたい



オープンウェイトの上位
Kimi K3 が指数57で4位

特定モデルの推奨ではありません。指数と単価は2026年7月28日時点のArtificial Analysis、解決率はOpenAI掲載値です。

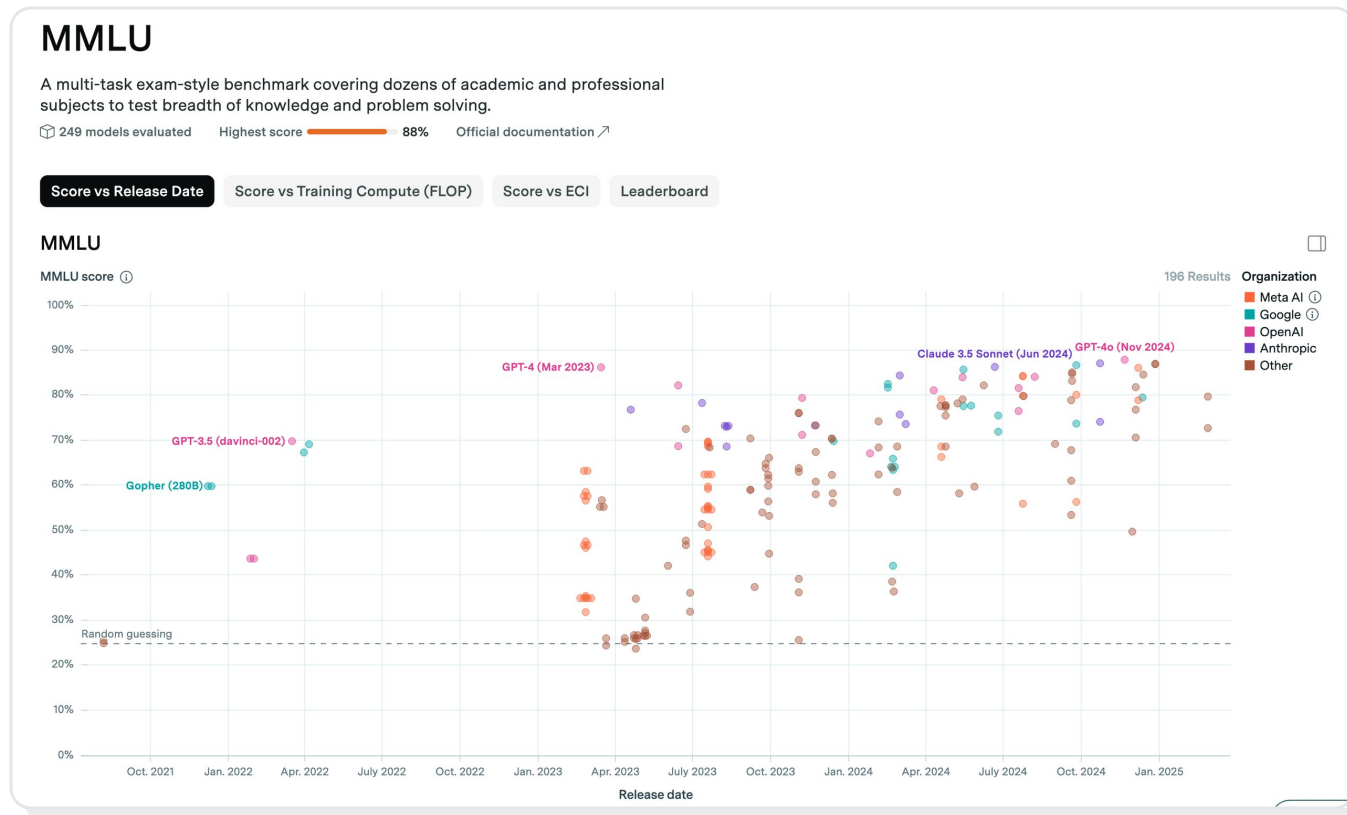
05

ベンチマークの世代交代

試験の正答率では差が出なくなり、比較の軸が費用と時間に移りました

従来ベンチマークの飽和

9割前後で頭打ちになった試験では**差が見えません**。



MMLU 最高88%で更新停止

評価済みは249モデル、リリース日の軸は2025年で止まり、2026年のモデルは登録されていません。

GPQA Diamond も同じ

上位8モデルが91～95%に密集し、モデル選定の判断材料になりません。

出典 Epoch AI 「MMLU」 (2026年7月28日アクセス)

現行世代のベンチマーク

測るものが正答率から**仕事の成果**へ移りました。

ベンチマーク	何を測るか	最高スコア（2026年7月時点）
ARC-AGI-2	見たことのない規則の発見	92.5%（GPT-5.6 Sol Max）
ARC-AGI-3	ルール非開示の環境を探索	30.2%（Claude Opus 5 High）
SWE-bench Verified	既存リポジトリのバグ修正	76.80%（同一ハーネス集計）
SWE-bench Pro	公開731問・平均107行の改修	61.5%（Muse Spark 1.1）
GDPval-AA v2	44職種の実務成果物を比較	1862 Elo（人間専門家＝1000）

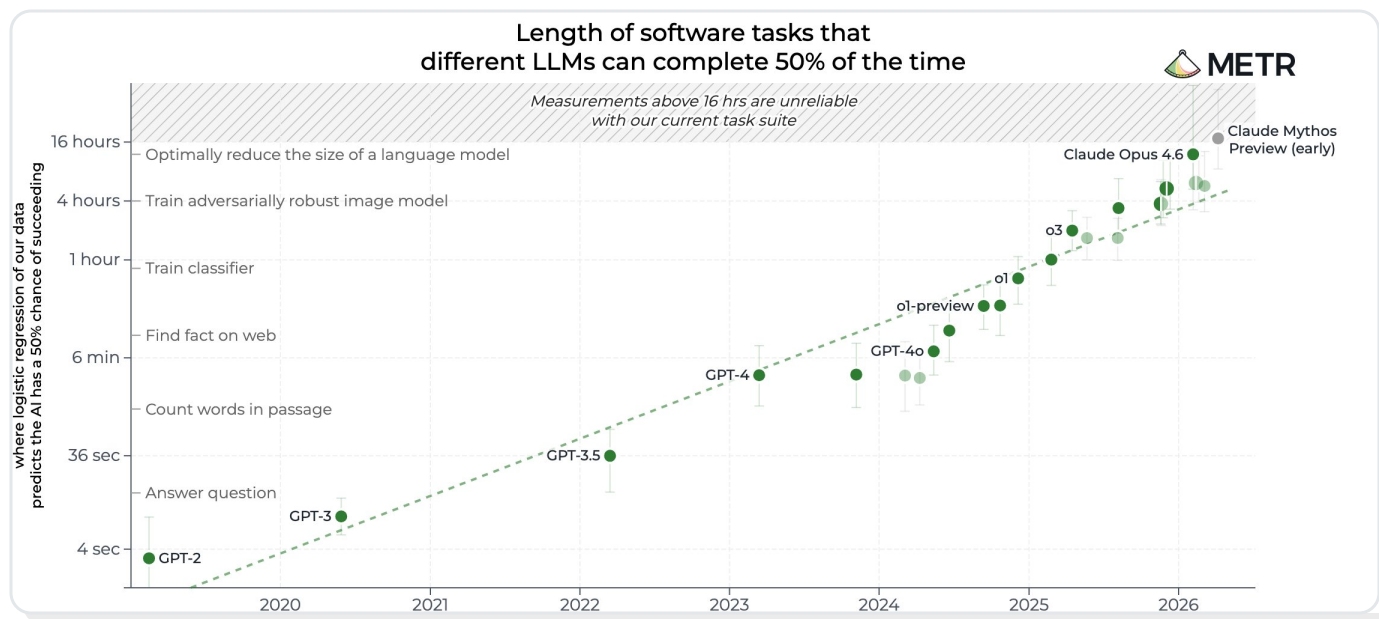
同じ名前でも別の数字

SWE-benchは公式集計の76.80%とベンダー自己申告の90%台が食い違います。順位は集計元をそろえて読みます。

出典 arcprize.org／swebench.com／labs.scale.com／artificialanalysis.ai（いずれも2026年7月時点）

人間何時間分の仕事かで測る

任せられる仕事の長さが**89日で倍**になります。



測り方

専門家が何分かかる課題まで、成功率50%で終わられるかを見ます。

課題を拡張

170問から228問へ増やし、8時間超の長課題を14問から31問にしました。

但し書き

METRは16時間を超える測定を現在の課題セットでは信頼できないと注記しています。倍化時間89日はTime Horizon 1.1（2026年1月29日）の推定値で、2019年以降で見ると196.5日です。

評価の揺れと自社での測定

公開ベンチの順位は**自社の判断材料になりません。**

◆ SWE-bench は答えが問題文に載っていた

成功パッチの32.67%は課題文やコメントに答えがあり、テストが弱いだけの見かけの成功も31.08%。除去すると解決率は12.47%から3.97%へ落ちました（SWE-Bench+、2024年10月）

◆ Terminal-Bench はベンチ側の修正で動いた

2.0から2.1で89問中28問を直した結果、モデルは同じままClaude Code + Opus 4.6が12.1ポイント上がりました（tbench.ai、2026年5月6日）

◆ FrontierMath は問題そのものが誤っていた

v2で全問題の42%に誤りが見つかり、12問削除・135問訂正（Epoch AI、2026年6月12日）

自社の案件で測った結果だけが判断材料になります。

実務での置き換え

公開ベンチの順位は候補を絞るまでです。自社の実データで小さな評価セットを作り、案件ごとに測ります。

06

国内企業の活用事例

社名と数値が公表されている事例だけを並べます

公表数値がある国内事例

公表されている数値は、どれも**特定業務の処理時間**です。

会社	対象業務	公表された効果	公表時点
パナソニック コネクト	図面・設計仕様の照合	1件50～340分が10分に（80～97%削減）	2026年2月
日本生命	法人向け書類の審査	1件27分が約8分に。年1万1000時間の見込み	2026年7月
三菱UFJ銀行	提案資料の作成	最大9割削減。行内2500名、年20万時間が目標	2026年2月
LINEヤフー	人事総務の10ツール	月1,600時間以上の削減見込み	2026年2月
ライフネット生命	社内用LLMの全社利用	利用率87%。2カ月で計約152時間削減	2025年4月

全社での削減時間まで公表しているのは1社だけです。

パナソニック コネクトは、AIアシスタント ConnectAI による業務時間削減を2024年度で年間約45万時間、前年比2.4倍と公表しています。

出典 各社プレスリリースおよび日本経済新聞（2026年7月16日）

パナソニック コネクトの図面照合

図面と設計仕様の照合が、50～340分から**10分**になりました。

効果：1回の作業時間の**80%～97%削減**を実現



パナソニック ニュースルームの発表（2026年2月19日）

対象業務

設計・開発部門の規格照合と外装部品の照合

仕組み

PDF図面から文字を抽出し、材質や仕上
げを文書間で自動照合します

基盤

Snowflake の Cortex AI

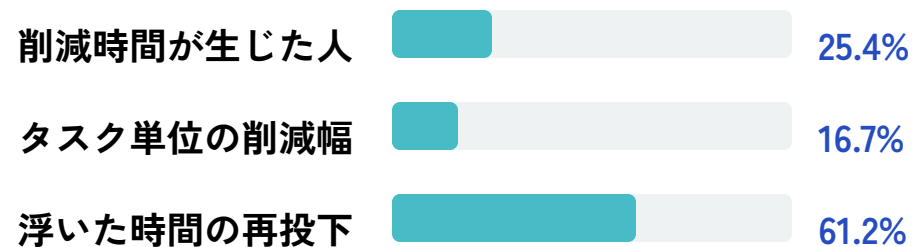
公表

2026年2月19日に社内利用を開始。対象
人数は非公表です

出典 news.panasonic.com

削減時間はどこへ行ったか

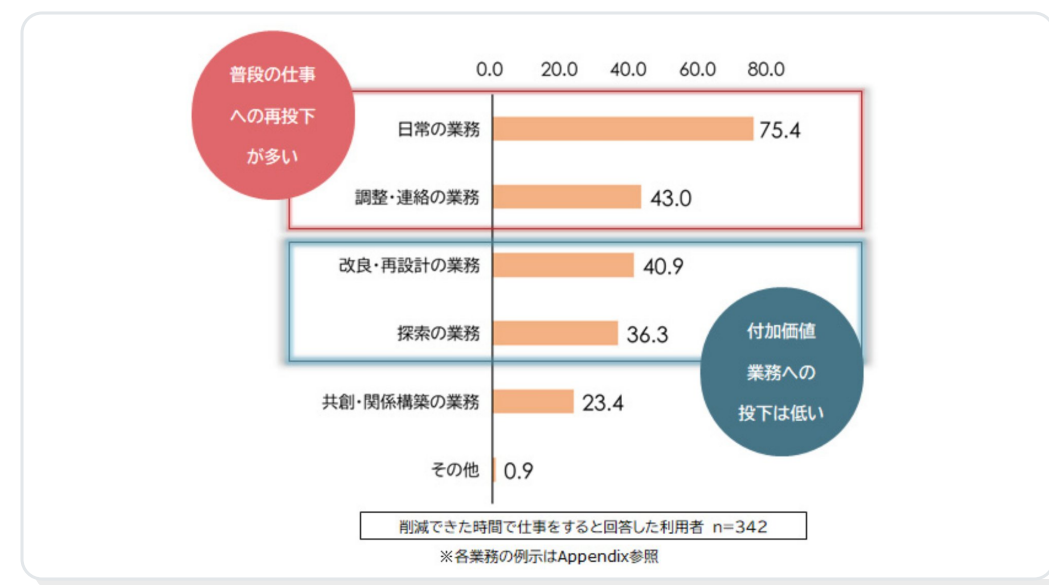
タスクは速くなっても、**全社の労働時間**はまだ減っていません。



削減が生じたと答えたのは25.4%で、削減幅はタスクレベルで平均16.7%、週26.4分にあたります。

浮いた時間を仕事に再投下した人は61.2%で、その再投下先は75.4%が日常業務でした。

全社の労働時間が減ったと示した公表事例は見当たりません。



パーソル総合研究所の調査（2026年3月13日）

参考 活用している企業は34.5%
（帝国データバンク、2026年5月、
有効回答1万312社）

07

勢力図と国家

モデルの性能よりも、電力と相互出資で並びが決まっています

計算資源の規模

競争の単位が、金額から**ギガワット**に変わりました。

9,650 億ドル

Anthropic の評価額

13 GW

Anthropic の計算資源

7,500 億ドル

OpenAI の支出計画

Stargate（OpenAI）の米国拠点 上位3件

拠点	稼働	計画
Abilene, Texas	0.3GW	1.2GW
Doña Ana County, NM	0	2.2GW
Shackelford County, TX	0	2.0GW

合計13GWの内訳

Google 5GW、AWS 5GW、AMD 2GW、Azure と NVIDIA で最大1GW。

売上はラン・レート470億ドル

2025年末は約90億ドル、2026年4月で300億ドル。
評価額は2026年5月28日の調達後の数字です。

全7拠点の計画は合計9GW超。出典 epoch.ai（2026年4月17日）、anthropic.com、CNBC、WSJ（2026年7月22日）

米国政府とモデル選定

政府がモデルの選定に、**直接介入**し始めました。



ホワイトハウスの発表（2026年7月22日）

同じ政権が、選定と排除を同時に進めています。

出典 whitehouse.gov（2026年7月22日）、CNBC（2026年3月5日）

Genesis Mission

15を超える連邦機関から50億ドル超を集め、278件を選定しました。

同じ政権が調達を止めた

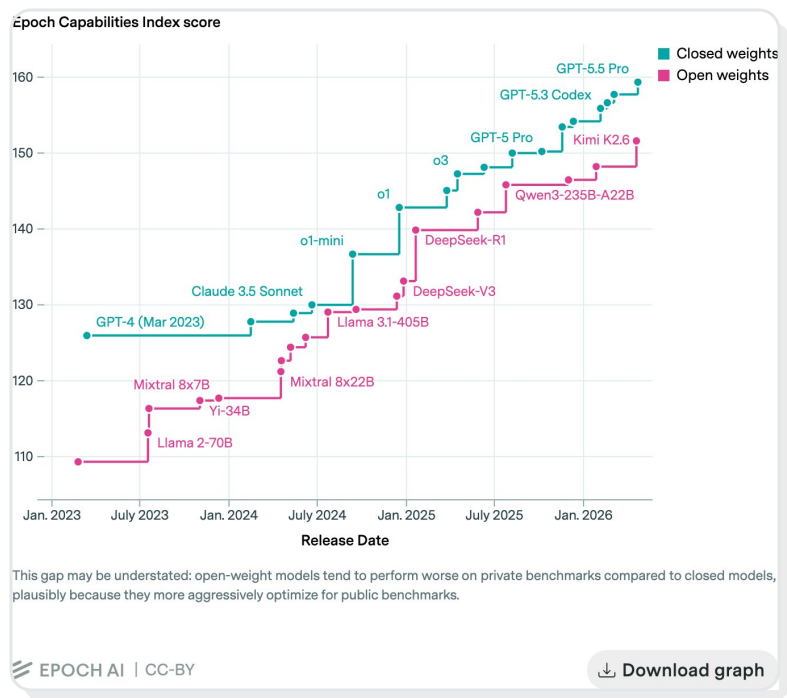
2026年2月27日に大統領が使用停止を指示し、3月5日に国防長官がサプライチェーンリスクへ指定。3月9日にAnthropicが提訴しました。

断ったのはAnthropic側

国防省が求めた無制限の利用を、大規模監視と完全自律兵器を理由に拒否しています。

中国勢とオープンウェイト

オープンウェイトとの差は、4か月まで縮みました。



上の線が重み非公開、下が重み公開

出典 epoch.ai (2026年5月28日)

差は4か月、ECI 8ポイント

2023年1月から2025年10月までは約3か月でした。2026年1月以降は小幅に広がっています。

出力単価は3.48ドル

DeepSeek V4-Pro は100万トークンあたり3.48ドルです。OpenAI は30ドル、Anthropic は25ドルでした。

25%の関税

米国は2026年1月15日から対中AIチップに25%の関税を課しています。

「オープンウェイトの禁止は主張していない」。求めるのは半導体の輸出規制強化、蒸留の取り締まり、共通の安全性テストの3点です。

Anthropic CEO Dario Amodei (2026年7月27日)

日本とEUの制度

日本は**促進**に寄り、EUは期日を後ろへ動かしました。

EU AI Act

- ・ 透明性義務（第50条）は2026年8月2日に適用
- ・ 高リスク（Annex III）は2027年12月2日へ
- ・ 組込型（Annex I）は2028年8月2日へ
- ・ 2026年5月6日にDigital Omnibusで暫定合意

人工知能基本計画（日本）

- ・ 2025年12月23日に閣議決定
- ・ 副題は「信頼できるAI」による日本再起
- ・ AI法 第18条第1項に基づく計画
- ・ 目標はAIを最も開発・活用しやすい国

規制の強弱より、いつ何が発効するかで社内の準備が決まります。

出典 cao.go.jp（人工知能基本計画）、Gibson Dunn（EU Digital Omnibus）

明日から見ておくもの

見るのは順位表ではなく、**自社の1タスク**です。

◆ 自社の業務で測る

自社の実データで同じ1業務を通し、時間と手戻りを記録します。公開ベンチの順位は、他社の課題で測った結果です。

◆ 1タスクいくらで見る

比べるのは1件あたりの費用です。単価が安いモデルでも、遅くて再試行が増えれば逆転します。

◆ 3か月で入れ替わる前提で作る

モデル名を手順書やコードに埋め込まず、差し替えられる形にしておきます。

測り方を決めておく

評価の記録先と担当を先に決めておくと、次のモデルが出たときの比較が1日で終わります。

出典一覧

数字の出どころは、[この一覧](#)からたどれます。

ドメイン	何の出典か
artificialanalysis.ai	知能指数、出力速度、ブレンダー単価、GDPval-AA
epoch.ai	オープン対クロードのECI差、Stargateの拠点容量
metr.org、arcprize.org、swebench.com	時間ホライズン、ARC-AGI、SWE-bench Verified
anthropic.com、claude.com、kimi.com	Claude Design、計算資源の提携、Kimi K3 の料金
gdconf.com、newsroom-deezer.com	ゲーム開発者調査と音楽配信の投稿状況
news.panasonic.com、 rc.persol-group.co.jp	国内事例の削減時間、生成AIとはたらしき方の実態調査
whitehouse.gov、cao.go.jp	Genesis Mission、人工知能基本計画

いずれも2026年7月28日までに公開された情報です。当日は数字が動いている場合があります。

