

Vals RSI Index (Vals 再帰的自己改善インデックス)

詳細・現状・今後の調査

2026 年 10 月 10 日時

Claude Opus 5.5

本文中の上付き番号は文末の参考文献に対応する。「【分析】」を付した記述は筆者の分析・推測であり、それ以外は出典で確認できる事実である。参考文献には、本文書作成時に原文を確認したか否かを付記した。

要点

- ・ Vals RSI Index は、米 AI 評価企業 Vals AI が 2026 年 8 月 13 日に公開した外部ベンチマークである。「AI が次のモデルを生み出す研究を、どこまで自律的にできるか」を測る^[1,2]。
- ・ 最新版（2026 年 10 月 1 日更新、4 タスク・24 モデル）の首位は Claude Opus 5.5（37.31%）で、Claude Fable 5.1（36.09%）、Claude Opus 5（33.02%）が続く^[1]。
- ・ 参照値（スコア 0.5＝既知の最良結果と同等）を超えたのは、Opus 5.5 の LM 訓練タスク 1 件だけである。事後学習タスクでは 24 モデル中 18 が 0 点だった^[1]。
- ・ Vals は、Anthropic 系列の指数が年 0.30 ポイントのペースで上昇しており、2027 年 7 月に「フロンティア級 AI 研究者」の目安とする 0.6 に達すると予測している（95%区間：2027 年 3 月～2028 年 5 月）^[1,3]。
- ・ 【分析】 この予測は、各モデル 1 回ずつの実行で得た少数のデータ点を直線で外挿したものである。タスク構成も公開後約 1 か月で 5 から 4 に変わった。到達時期の予測ではなく、上昇トレンドを示す早期警戒シグナルとして扱うのが妥当である。

1. 指標の概要

1.1 発表主体と公開の経緯

Vals AI は 2024 年創業のサンフランシスコ拠点の AI 評価企業で、共同創業者の Rayan Krishnan 氏（25 歳、スタンフォード大学出身）が率いる^[4]。法務・金融・コーディングなどの業務別に非公開テストセットを用意してモデルを評価し、評価料を AI 企業から受け取る事業である^[4]。Krishnan 氏はこの収益モデルを、受験生が SAT の受験料を College Board に払う構図に例えている^[4]。

2026 年 8 月 13 日、Vals は a16z 主導の 4,000 万ドルの Series A（評価額 4 億ドル）を発表した。同

時に「Frontier Risk Benchmarks」の一つとして RSI Index を公開し、CoreWeave との協業によるものとした^[2]。同社は、売上が 2025 年通年比で 8 倍、顧客数が倍増、チームが 6 か月で 3 倍になったことを挙げる。また、OpenAI・Anthropic・Google・Meta・xAI のモデルカードで自社の評価結果が引用されたとしている。いずれも自社発表である^[2]。

公式ページの著者は Oliver Chen、Geby Jaff、Langston Nashold、Rayan Krishnan の 4 氏である。計算資源は CoreWeave、実験記録用のノートブックは Marimo が提供している^[1]。論文版は 2026 年 9 月 9 日付で alphaXiv に初版が投稿され^[5]、2026 年 10 月付の改訂版も公開されている^[3]。arXiv 本体への掲載は確認できなかった。

1.2 目的と RSI の位置づけ

公式ページは、I.J. Good (1965) が提唱した「より優れた機械を設計できる超知能機械」を起点に、RSI を「人間の関与なしに後継モデルを自律開発すること」と位置付ける^[1]。RSI は単一の能力ではなく、アーキテクチャ、アルゴリズム、データの取得と選別、評価設計、システム工学の総体だとしている^[1]。

RSI Index が測るのは、このうち LLM 開発の研究タスクの一部を自律的に遂行できる度合いである。論文自身も、エージェントに後継モデルを訓練させ、さらにその後継を訓練させる「ループの完結」は検証していないと明記している^[3]。

1.3 評価方法（現行 4 タスク）

各タスクは、計算資源と時間を固定した自由度の高い研究課題である。エージェントは自ら実験して成果物を 1 つ提出し、成果物は別のサンドボックスで独立に検証される^[1,3]。モデルは各社の標準エージェント環境で、最大の推論努力設定で動かす。Claude と GLM は Claude Code、GPT は Codex、Gemini は Antigravity、Kimi は Kimi Code、Qwen は Qwen Code、Muse は Muse Code、Grok は Grok Build を使う^[1]。エージェントが予算を残して作業を終えた場合は、監督プロセスが作業空間を引き継いで新しいセッションを起動する。追加起動の回数は中央値で 10 回だった^[3]。

表 1 現行 4 タスクの概要^[1,3]

タスク	内容	資源・時間	0 (基準)	0.5 (参照値)	0.6 (目標)
圧縮	enwik8 (英語版 Wikipedia 100MB) の可逆圧縮。復元器のサイズも含めて計測	24 時間、H100×1、オフライン	gzip 2.92 BPB	NNCP 1.2435 BPB	1.070 BPB

タスク	内容	資源・時間	0（基準）	0.5（参照値）	0.6（目標）
LM 訓練	小型バイトレベル GPT（4 層・幅 256）の訓練プログラム改善。提出後に 3 シードで 30 分ずつ再訓練し、非公開テストで評価	24 時間、H100×1	1.3699 BPB	Al-Rfou ら（2019）0.85 BPB	0.789 BPB
ハーネス設計	固定モデル（Gemini 3.5 Flash-Lite）を審査員とする判定ハーネスの改良。非公開の 24 アプリ・205 ステップで評価	12 時間、GPU なし	50%	人間アノテーター間の最高一致率 93.6%	95.6%
事後学習	Qwen3.6-35B-A3B を Finance Agent v2 向けに改善。非公開 126 問×3 回で評価	30 時間、RTX PRO 6000×16、ネット接続可、外部 API \$100	35.38%	Muse Spark 1.2 59.42%	71.8%

注：BPB=bits per byte（小さいほど良い）。ハーネス設計と事後学習は大きいほど良い。

1.4 採点方式

各タスクの結果は、基準値（0 点）と理論上の最良値（1 点）の間で、対数スケールにより「縮めた差の割合」として正規化する。さらにタスクごとの指数 γ で変換し、参照値がちょうど 0.5 点になるよう調整する^[3]。 γ は 0.32（事後学習）から 1.26（LM 訓練）の範囲で、この変換は MLS-Bench の方式を踏襲している^[3]。基準を下回る結果や無効な結果は 0 点とし、指数は 4 タスクの平均である^[3]。

0.6 は「既知の最良結果を明確に上回る」水準として設定されている。Vals は歴史的な技術的飛躍を当てはめてこの水準を校正した。BERT（MNLI の誤り率低減）なら 0.58、AlexNet（ImageNet top-5 誤り率 37%減）なら 0.62、LSTM から Transformer への移行（text8）なら 0.63 に相当するという^[3]。

なお、参照値には過去の研究結果、人間アノテーターの一致率、フロンティアモデルのスコアが混在している。Vals 自身も、これは最強の人間性能を統一的に測る尺度ではないと認めている^[1]。同一条件下で人間専門家と比較したベースラインはない^[5]。

1.5 版の変遷

9 月 9 日の初版と 10 月の改訂版では構成が大きく異なる。数値を引用・比較する際は、どちらの版かを確認する必要がある。

表 2 初版と現行版の比較^[1,3,5]

項目	初版（2026 年 9 月 9 日）	現行版（2026 年 10 月）
タスク数	5（Parameter Golf を含む）	4（Parameter Golf を除外）
評価モデル数	8	論文 19、公式ページ 24

項目	初版（2026 年 9 月 9 日）	現行版（2026 年 10 月）
首位	Claude Fable 5.1（0.350）	Claude Opus 5.5（0.373）
ハーネス設計の参照値	95%	93.6%（人間アノテーター間の最高一致率）
参照値（0.5）の超過	なし	Opus 5.5 の LM 訓練のみ

注：Parameter Golf は、16MB 以内のモデルを 8×H100 で 10 分間訓練し、FineWeb で評価するタスク^[5]。除外理由を説明した公開資料は確認できなかった。

1.6 類似評価との違い

表 3 主な類似評価との比較

評価	主体	測るもの	RSI Index との違い
RE-Bench	METR	7 つの ML 研究環境で、エージェントと人間専門家を同じ 8 時間予算で比較	RSI Index は人間との同条件比較ではなく、はるかに大きな資源で得られた「既知の最良結果」を基準とする ^[3]
PaperBench／MLE-bench	OpenAI	ICML 論文 20 本の再現／Kaggle 型 ML 工学 75 課題	機械学習全般が対象。RSI Index は LLM 開発工程に絞り、リリースや提供元をまたいで推移を追う ^[3]
Time Horizons ほか	METR ほか	長時間タスクの自律遂行（主にソフトウェア工学）	長期的な信頼性の測定が主眼で、AI 研究開発の有効性は測らない ^[3]
OpenAI 社内の「RSI Index」	OpenAI	社内の AI 研究自動化に関する集約指標（システムカードで報告） ^[6]	名称は同じだが別物。GPT-5.6 Sol で 57.9、GPT-5.5 で 41.7 とされる（二次情報） ^[7,8]
R&D 自動化の社内測定	Anthropic Institute	社内のモデル研究開発タスクのうち、Claude が「主導」する割合	実業務に基づく社内測定であり、外部ベンチマークではない ^[3,9]

注：Vals 論文は、AI 研究能力の測定がこれまで各開発企業の社内タスクに依存し、企業間の比較や独立した追跡ができなかった点を、本指標の存在意義としている^[3]。

2. 現状

2.1 最新リーダーボード

表 4 公式リーダーボード（2026 年 10 月 1 日更新、24 モデル）の上位と主要モデル^[1]

順位	モデル	指数	圧縮	LM 訓練	ハーネス	事後学習
1	Claude Opus 5.5	37.31%	0.465	0.530	0.145	0.351
2	Claude Fable 5.1	36.09%	0.443	0.495	0.226	0.280
3	Claude Opus 5	33.02%	0.426	0.428	0.148	0.318
4	Gemini 4 Argon	30.55%	0.447	0.480	0.139	0.154
5	GPT-6 Sol	28.1%＊	0.423	0.311	0.228	0.163
6	GPT-6 Astra	27.1%	0.426	0.484	0.095	0.078

順位	モデル	指数	圧縮	LM 訓練	ハーネス	事後学習
7	GPT-6.1 Sol	26.6% *	0.435	0.429	0.201	0.000
8	Grok 4.6	25.5% *	0.378	0.423	0.217	0.000
11	GPT-5.6 Sol	23.9%	0.410	0.361	0.185	0.000
12	GLM 5.3	22.9% *	0.334	0.408	0.174	0.000
13	Qwen 3.8 Max	21.8%	0.353	0.383	0.135	0.000
14	Kimi K3	20.9%	0.333	0.404	0.099	0.000
16	Gemini 3.8 Flash	19.9%	0.290	0.362	0.145	0.000
17	Muse Spark 1.3 Max	19.6% *	0.231	0.409	0.145	0.000
24	Kimi K2.5	6.8% *	0.000	0.253	0.021	0.000

注：* は公式ページのタスク別スコアから筆者が算出した 4 タスク平均、順位も同様に算出（公式ページ掲載の順位と整合）。表にない 9～10 位は Claude Fable 5 と Claude Opus 4.8 で、両者の圧縮・LM 訓練の結果は同一値である。論文によれば、Fable 5 がこの 2 タスクを拒否したため、Opus 4.8 で代替した^[3]。DeepSeek のモデルは評価対象に含まれていない^[1]。

2.2 タスク別の到達度

- ・ **圧縮**：首位 Opus 5.5 の 1.312 BPB は、参照値 1.2435 BPB に届いていない。上位の提出物は、H100 を使えるにもかかわらず、ニューラル圧縮ではなく適応的コンテキストミキシングを選んだ^[1]。
- ・ **LM 訓練**：全モデルが基準を改善した。Opus 5.5 の 0.8308 BPB は参照値 0.85 BPB を下回り、全タスクを通じて参照値を超えた唯一の例である^[1]。ただし、参照論文と同じ訓練・評価条件での再現ではない。参照論文は TPU v2 で 175 時間訓練したのに対し、Opus 5.5 のレシピは単一 H100 で 30 分の訓練である^[1]。
- ・ **ハーネス設計**：首位は GPT-6 Sol の 81.46% で、Fable 5.1 の 81.30% が僅差で続く^[1]。人間の一致率 93.6% との差はまだ大きい。なお、論文の 19 モデルの範囲では Fable 5.1 が首位である^[3]。
- ・ **事後学習**：首位 Opus 5.5 の 44.53% に対し、親モデルは 35.38% である^[1]。親モデルを標準誤差の 2 倍を超えて上回ったのは、Opus 5.5、Fable 5.1、Opus 5 の 3 モデルだけだった^[3]。Fable 5.1 は訓練したアダプタを 2 つとも棄却し、親モデルの重みに金融向けプロンプトを組み合わせで提出した^[1]。

2.3 推移と Vals の予測

Anthropic 系の指数は、7 リリースで 0.148（Opus 4.5、2025 年 11 月）から 0.373（Opus 5.5、2026 年 9 月）へ単調に上昇した。OpenAI 系は 5 リリースで 0.158 から 0.271、Moonshot 系は 0.068 から 0.209 である^[3]。

表 5 Vals による線形外挿と感度分析^[3]

項目	Anthropic	OpenAI
傾き（年あたりの指数上昇）	0.30（95%区間 0.17～0.42）	0.17
0.5 到達	2027 年 3 月（2026 年 12 月～2027 年 10 月）	2028 年 2 月
0.6 到達	2027 年 7 月（2027 年 3 月～2028 年 5 月）	2028 年 9 月
0.6 到達（1 リリース除外）	2027 年 3 月～8 月	2027 年 12 月～2029 年 5 月
0.6 到達（1 タスク除外）	2027 年 3 月～11 月	2028 年 1 月～2029 年 5 月

注：区間は Anthropic のみ算出。OpenAI は 5 リリースしかなく、傾きの不確実性が大きいと Vals 自身が注記している^[3]。

Vals はこの予測を各社の社内指標と照合している。Anthropic の発表によれば、Claude が「主導」する社内のモデル研究開発タスクの割合は、2026 年 3 月の 1%から 8 月には 26%に上がった。ロジスティック当てはめでは 2027 年 8 月、線形当てはめでは 2027 年 10 月に 99%へ達する^[3,9]。OpenAI は、2028 年 3 月までの「自動 AI 研究者」の実現に向けて強い進展があると報告している^[3,10]。Vals は、Anthropic の社内予測が自社の推計と 3 か月以内の差に収まると主張する^[3]。Vals は、Opus 5.5 の結果を受けて予測時期を前倒ししたことも X 上で発信している^[11]。

【分析】これらの社内指標はいずれも各社の自己測定であり、互いに独立した検証ではない。数字が整合していても、予測の確度が高まるとは限らない。

2.4 エージェントの研究行動に関する Vals の所見

- ・ **実行力が判断力を上回る**：好成績は既存技術（RoPE、RMSNorm、SwiGLU、LoRA など）の組み合わせと調整から生まれた。根本的に新しい手法は見られない^[1,3]。
- ・ **実験量は成績を予測しない**：GPT-5.6 Sol は LM 訓練で 114 回実験したが、76 回の Opus 5 より成績が悪かった^[3]。
- ・ **事後学習の手法が狭い**：訓練済みの提出物はすべて教師ありファインチューニング（ほぼ LoRA/QLoRA）によるものだった。DPO を使ったのは Grok 4.6 だけで、結果は親モデルを下回った^[3]。
- ・ **改善を積み重ねられない**：Opus 5 に金融、次いで法務の順で事後学習させたところ、法務の基準達成率は 72.64%から 67.29%に、金融のスコアは 42.20%から 38.55%に下がった^[3]。
- ・ **並列化は効果がなかった**：GPT-5.6 Sol のマルチエージェントモードは、ハーネス設計で 69.92%（API 費用 751 ドル）にとどまり、単一エージェント（77.89%、259 ドル）を下回った^[3]。

2.5 反響・評価

TechCrunch は Vals の事業全体を紹介する中で、RSI ベンチマークへの取り組みに触れた^[4]。AlphaSignal は、Vals の予測が制約付きベンチマークからの外挿であり、無制限の自己改善を直接実証するものではないと注記している^[12]。

一方、LessWrong、Epoch AI、METR 関係者などによる体系的な論評は、調査の範囲では見つからなかった。主要ラボのシステムカードが Vals RSI Index を引用した事実も確認できなかった。Vals の「モデルカードで引用された」という主張は、同社の評価結果全般についてのものである^[2]。

3. 方法論上の限界

Vals 自身が論文で認めている限界と、筆者が加える指摘を以下に整理する。

1. **単発実行**：各モデル・タスクの組み合わせは1回しか実行されていない。誤差棒は凍結した成果物の再評価のばらつきにすぎず、実行ごとの分散は反映していない^[3]。
2. **ハーネスとの交絡**：各社の標準環境で評価するため、モデルの能力とツールの品質を切り分けられない^[3]。
3. **データ汚染**：圧縮と LM 訓練は公開データと公開結果に基づくため、汚染の可能性を排除できない^[3]。
4. **タスクの狭さ**：データキュレーション、カーネル最適化、安全性研究、大規模システム工学は対象外である^[5]。
5. **ループ未完結・一般化未検証**：後継モデルの訓練を連鎖させる検証は行っておらず、フロンティア規模の開発へ一般化できるかも直接は検証していない^[3]。
6. **【分析】構成変更**：公開後約1か月で5タスクから4タスクに変わり、時系列の連続性が損なわれた。除外の理由も公表されていない。
7. **【分析】利益相反の可能性**：Valsはラボから評価料を受け取る事業者であり^[4]、劇的な予測は資金調達・広報と同時期に出されている^[2]。不正を示すものではないが、第三者による再現の必要性を高める要因である。

4. 今後

4.1 Vals のロードマップ

Vals は今後も新しいフロンティアモデルを評価し、トレンドが続くかを追跡するとしている^[3]。将来版の改善として、タスクごとの複数回実行と、データキュレーション、カーネル最適化、継続的事後学習などのタスク追加を挙げる^[3]。具体的な日程は示されていない。

4.2 RSI・AI 安全性をめぐる議論への含意

- ・ **社内判定との乖離**：Anthropic は Claude Opus 5 のシステムカードで、自動 AI 研究開発の能力閾値には達していないと判断したとされ^[13]、Opus 5.5 についても同様と報じられている^[14]。
- ・ **【分析】外部指標は「1 年以内に既知の最良結果を超える」と示唆する一方、ラボの公式判定は「閾値未満」である**。この乖離は、閾値の定義（研究者全体を代替できる水準）と、部分的なタスクでの性能との差を反映している。
- ・ **【分析】外部観測の価値**：本指標は、規制当局や第三者がラボの外から継続的に観測できる早期警戒シグナルとして価値がある。ただし、ラボの閾値判定に代わる根拠にはならない。
- ・ **【分析】知能爆発論への示唆**：実験の実行と反復は急速に自動化されつつある。一方で、研究方向の選択、新手法の発明、改善の積み重ねは弱い。知能爆発の律速段階が「研究判断」と「改善の積み重ね」にあるという見方と整合的であり、現時点のデータは急激な離陸を支持していない。

4.3 見通し

【分析】 以下はいずれも筆者の推測である。

- ・ **短期（～2026 年末）**：首位スコアは 0.38～0.42 程度に上がる可能性が高い。年 0.30 の傾きと、2～3 か月ごとのフロンティアモデルのリリースが根拠である。伸びしろはハーネス設計と事後学習にある。
- ・ **0.5 到達（2027 年前半）**：確度は中程度である。LM 訓練は既に参照値を超え、圧縮も接近している。一方、事後学習（参照値 59.42% に対し最高 44.53%）とハーネス設計がボトルネックであり、指数の上昇はこの 2 タスクの改善に大きく依存する。
- ・ **0.6 到達（2027 年 7 月）**：懐疑的に見るべきである。少数のデータ点への線形外挿であることに加え、0.5 を超えるには公開結果の再現ではなく最良値の更新が必要になる。既存技術の組み合わせで伸びてきた現在のペースが続く保証はない。Vals 自身も、新手法が必要かどうかは未解決の問題としている^[3]。

- ・ **指標の持続性**：参照値を超えるタスクが増えれば、タスクの入れ替えや尺度の見直しが起き、時系列が再び途切れる可能性がある。

5. 引用・活用上の留意点

- ・ 数値を引用するときは、版と日付を併記する（例：「Vals RSI Index（2026 年 10 月 1 日更新、4 タスク版）で Claude Opus 5.5 が 37.31%」）。初版（5 タスク）の数値とは混在させない。
- ・ OpenAI 社内の同名指標「RSI Index」とは明確に区別する^[7,8]。
- ・ 「2027 年 7 月」は、Vals 社による Anthropic 系列の線形外挿（95%区間：2027 年 3 月～2028 年 5 月）として紹介し、RSI 到達時期のコンセンサスとしては扱わない。
- ・ 今後の注目点：Parameter Golf 除外理由の説明、複数回実行版の公開、DeepSeek など未評価モデルの追加、第三者による再現、主要ラボのシステムカードでの言及の有無。

参考文献

【確認済】本文書作成時（2026 年 10 月 10 日）に原文を閲覧し、記載内容を確認したもの。【未確認】原文を閲覧していないもの（URL・内容は引用元文献または二次情報による）。

- [1] Vals AI (Oliver Chen, Geby Jaff, Langston Nashold, Rayan Krishnan) 「Vals RSI Index Leaderboard and Methodology」 Vals AI, 2026 年 10 月 1 日更新. https://www.vals.ai/benchmarks/rsi_index 【確認済】
- [2] Rayan Krishnan 「Series A: Always a Higher Peak」 Vals AI Blog, 2026 年 8 月 13 日. <https://www.vals.ai/blogs/series-a> 【確認済】
- [3] Vals AI Team 「RSI-Index: Measuring Autonomous LLM R&D in Frontier Agent Systems」 Preprint（改訂版、4 タスク・19 モデル）, 2026 年 10 月. <https://www.alphaxiv.org/abs/2609.rsi-index-autonomous-llm-rd.pdf> 【確認済】
- [4] Lucas Ropek 「Vals, backed by Andreessen Horowitz, is looking to become the gold standard for AI benchmarking」 TechCrunch, 2026 年 9 月 19 日. <https://techcrunch.com/2026/09/19/vals-backed-by-andreessen-horowitz-is-looking-to-become-the-gold-standard-for-ai-benchmarking/> 【確認済】
- [5] Vals AI Team 「RSI-Index: Measuring Autonomous LLM R&D in Frontier Agent Systems」 alphaXiv（初版、5 タスク・8 モデル）, 2026 年 9 月 9 日投稿. <https://www.alphaxiv.org/abs/2609.rsi-index-autonomous-llm-rd> 【確認済】
- [6] OpenAI 「GPT-5.6 Sol System Card」 2026 年 7 月 9 日. <https://deploymentsafety.openai.com/gpt-5-6/gpt-5-6.pdf> 【未確認：文献[3]の参考文献から転記】
- [7] Tech Jack Solutions 「GPT-5.6 に関する解説記事」（標題・日付未確認）. <https://techjacksolutions.com/ai-tools/chatgpt/gpt-5-6-deep-dive/> 【未確認：二次情報】
- [8] LLM Stats 「RSI Index」 ベンチマーク掲載ページ（日付未確認）. <https://llm-stats.com/benchmarks/rsi-index>

【未確認：二次情報】

- [9] Anthropic Institute 「Measurements for Understanding the Pace of AI Development inside Frontier Labs」 2026 年 9 月. <https://www.anthropic.com/institute/measuring-pace-of-ai-development> 【未確認：文献[3]の参考文献から転記】
- [10] OpenAI 「Research Acceleration: The View inside OpenAI」 2026 年 9 月. <https://openai.com/index/research-acceleration-view-inside-openai/> 【未確認：文献[3]の参考文献から転記】
- [11] @ValsAI (Vals AI 公式 X アカウント) 投稿, 2026 年 9 月 (日付未確認). <https://x.com/ValsAI/status/2102445426222010795> 【未確認：原文未取得】
- [12] AlphaSignal 「Vals AI crowns Claude Opus 5.5 king while MiMo costs 100x less」 (日付未確認). <https://alphasignal.ai/news/vals-ai-crowns-claude-opus-5-5-king-while-mimo-costs-100x-less> 【未確認：原文未取得】
- [13] Anthropic 「Claude Opus 5 System Card」 2026 年 7 月 24 日. <https://www-cdn.anthropic.com/c5fbac3f0b1280a933ebd26d3cb8bb9f5bdeaf48/Claude%20Opus%205%20System%20Card.pdf> 【未確認：URL・公表日は文献[3]の参考文献で確認、本文で引用した判定内容は原文未確認】
- [14] @ZombieCodeKill 「Claude Opus 5.5 is here」 Medium (日付未確認). <https://medium.com/@ZombieCodeKill/claude-opus-5-5-is-here-8307a54dcf0b> 【未確認：二次情報】