◈ 意見

AGIがどうでもいい理由

AGIは汎用人工知能の略で、知的な課題の大半をこなせるとされる仕組みです。何をAGIと呼び、どうテストするかは研究所の間でも一致していません。私は現在のツールを毎日使っています。このページでは、実証できることと人々の予測を分けます。

Fabio Akita · ailair.akitaonrails.com/ja/agi/

*AI*への熱狂は、*AI*についての知識に反比例します。

Sua empolgação com IA é inversamente proporcional ao seu conhecimento sobre IA.

1. 定義

それが何なのか、誰も合意していない

5つの研究所にAGIの意味を聞けば、5つの違う的が返ってきます。そのうち1つはただの利益額です。



1. 定義

それが何なのか、誰も合意していない

- OpenAIの憲章では、AGIとは経済的に価値のある仕事の大半で人間を上回るシステムのことです。
- Google DeepMindの論文「Levels of AGI」は、2023年の時点でChatGPTを「Emerging AGI」に分類していました。フランソワ・ショレは知能を、新しいスキルをどれだけ効率よく習得できるかと定義しています。ほかの定義はどれもそれを測っていません。
- MicrosoftとOpenAIの契約では、AGIは1,000億ドルの利益と定義されていたと報じられています。認知に関する要素は何もありません。2026年4月、この条項はまるごと削除されました。
- 売る側でさえ肩をすくめています。サム・アルトマンは「あまり役に立つ用語ではない」と言い、ダリオ・アモデイはこの言葉が嫌いだと言い、グレッグ・ブロックマンは「AGIの時代」の幕開けを宣言しながら、定義は人によって違うと認めました。

2. テスト

だから誰も テストできない

定義できないものは測れません。代わりにあるのは、埋まっては攻略され、次のものに置き換えられるベンチマークの列です。



2. テスト

だから誰もテストできない

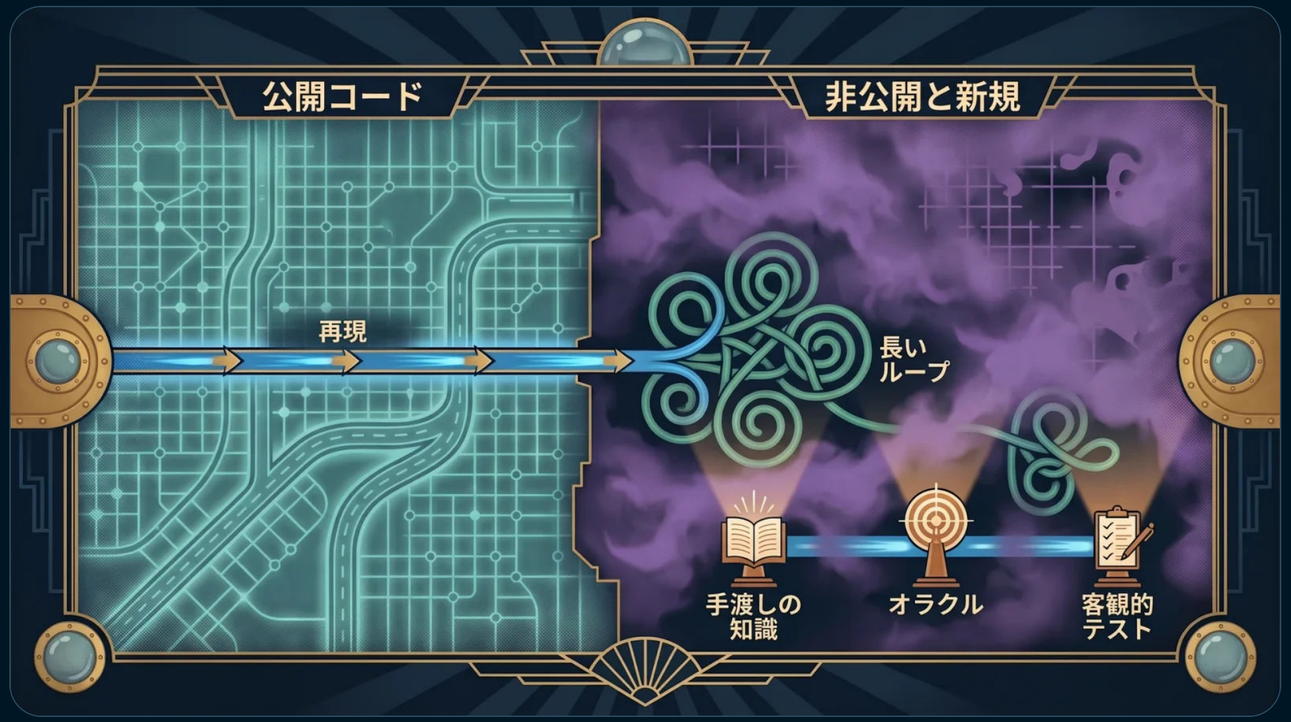
- ARC-AGI-3は2026年3月に公開され、最高のAIでも約0.5%でした。9月には専用のハーネスが19,000ドルで99.9%まで持っていきました。ARC Prizeのチーム自身が、これを飽和させてもAGIの証明にはならないと書いています。
- Humanity's Last Examは、生物と化学の解答の約29%が誤りの可能性があるまま公開されました。OpenAIはSWE-bench Verifiedの報告をやめました。テストに欠陥があり、モデルが解答を見てしまっていたからです。MetaはLMArenaに特別に調整したLlama 4を出しました。
- 同じモデルが「Definition of AGI」のテストで57%にも24%にもなります。数字の平均の取り方しだいです。
- 私自身のベンチマークも2回飽和し、隠れた妨害工作を軸に作り直しました。今では複数のモデルが100点で並んでいます。フロンティアは天井に達しつつあります。比べるものとして残っているのはコスト、速度、規律です。

> ***私たちが自分で何かのAGIのマイルストーンを達成したと主張するなんて、私に言わせればばかげたベンチマークハックにすぎません。***
>
> *サティア・ナデラ、Microsoft CEO*

◆ 3. フロンティア

あるものは得意、ないものは高くつく

どのモデルも公開されているものから学びました。問題がその地図の外に出ると、進み方は試行錯誤になり、請求額が一気に膨らみます。



◈ 3. フロンティア

あるものは得意、ないものは高くつく

0%

どのタスクも新しいARC-AGI-2での、純粋なLLMのスコア

- 似たものが学習データにあったなら、モデルはそれを速く安く再現します。それがソフトウェア、クローン、「Adobeは終わった」にとって何を意味するかは、別のページにまとめました。

- AlphaEvolveとo3は、安い検証器を使った探索で新しいものを見つけました。この方法は有効ですが、大規模にできるのは6社ほどしかありません。

1300億

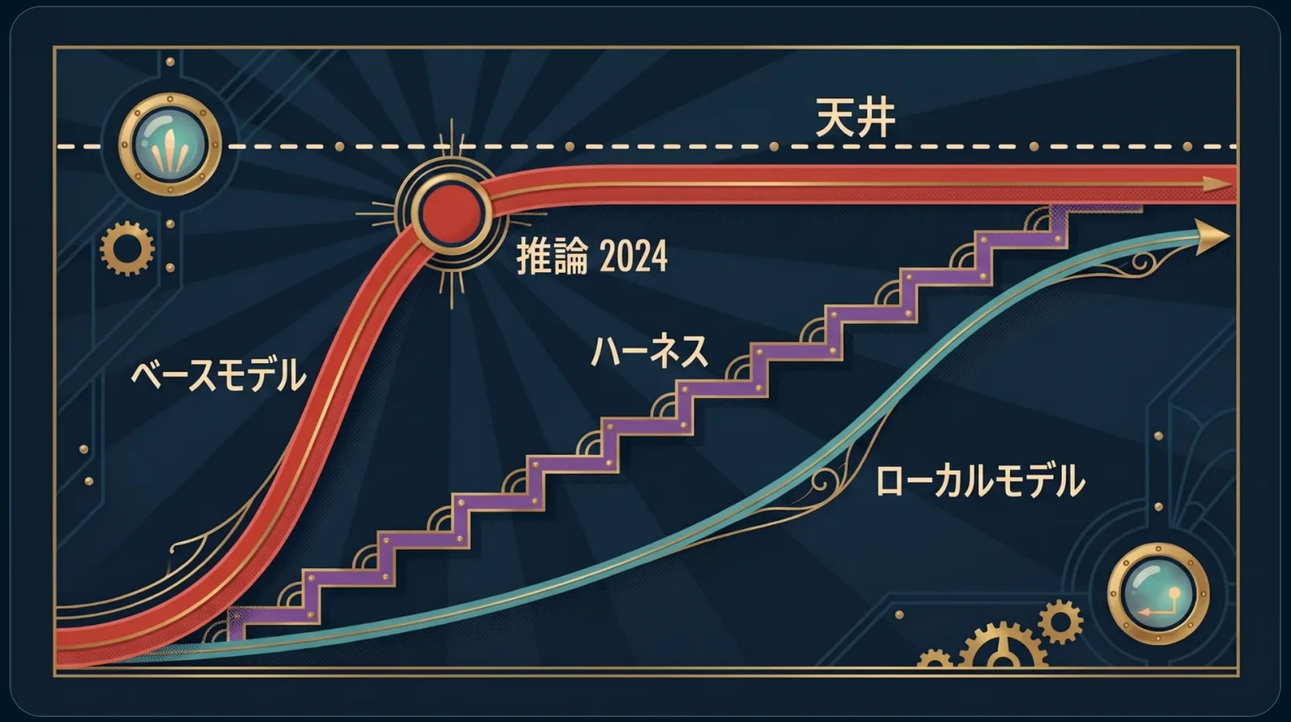
出力トークン。ナビエ・ストークスの証明1つで約1,500万ドル

- なかったなら、進み方は試し、後退し、元に戻す長いループになります。OpenAIのナビエ・ストークスの証明には、約1万のエージェントを88時間動かし、出力トークン約1300億、定価で約1500万ドルかかりました。それでも数十年分の公開された数学に頼っていました。

4. 頭打ち

モデルはあまり賢くならなくなった

私の見方では、素の知能で最後に大きく伸びたのは2024年の推論です。それ以降の飛躍に見えたものは、モデルの周りのツールから来ています。



4. 頭打ち

モデルはあまり賢くならなくなった

- 2025年は同じ知能を安くしました。Mixture of Experts、プロンプトキャッシュ、長いコンテキスト。2026年はそれを、検索し、テストを走らせ、推測を間違えたところを直すエージェントの中に入れました。
- スタンフォード大学のAI Indexによると、Chatbot Arenaの1位と10位のモデルの差は1年で11.9%から5.4%に縮みました。私のベンチマークでは、Opus 4.5以降、フロンティアモデルは同じ層に並んだままです。
- 2024年11月、Reutersは大手研究所で事前学習の効果が小さくなっていると報じました。その1か月後、イリヤ・サツケバーは、インターネットは1つしかないので「私たちが知っている形の」事前学習は終わると言いました。
- その天井の下で、小さなモデルが差を詰め続けています。Epoch AIによると、2026年のオープンウェイトモデルはフロンティアから約4か月遅れで、ゲーミングGPU1枚で6〜12か月前のフロンティアが動きます。私の見方では、今日のフロンティアはいずれ自分のマシンで動くようになります。

4. 頭打ち

単独のLLMは高価なオウムです。エージェント付きのLLMは、本当に仕事ができるくらい有能です。

LLM sozinho é um papagaio caro. LLM com agente é competente o suficiente pra trabalhar de verdade.

私、「LLM＋エージェント＝十分に有能」

5. 学習

仕事をしながら学ぶことはない

モデルは学習後に固定されたファイルです。データセンターはそのコピーを動かしています。



5. 学習

仕事をしながら学ぶことはない

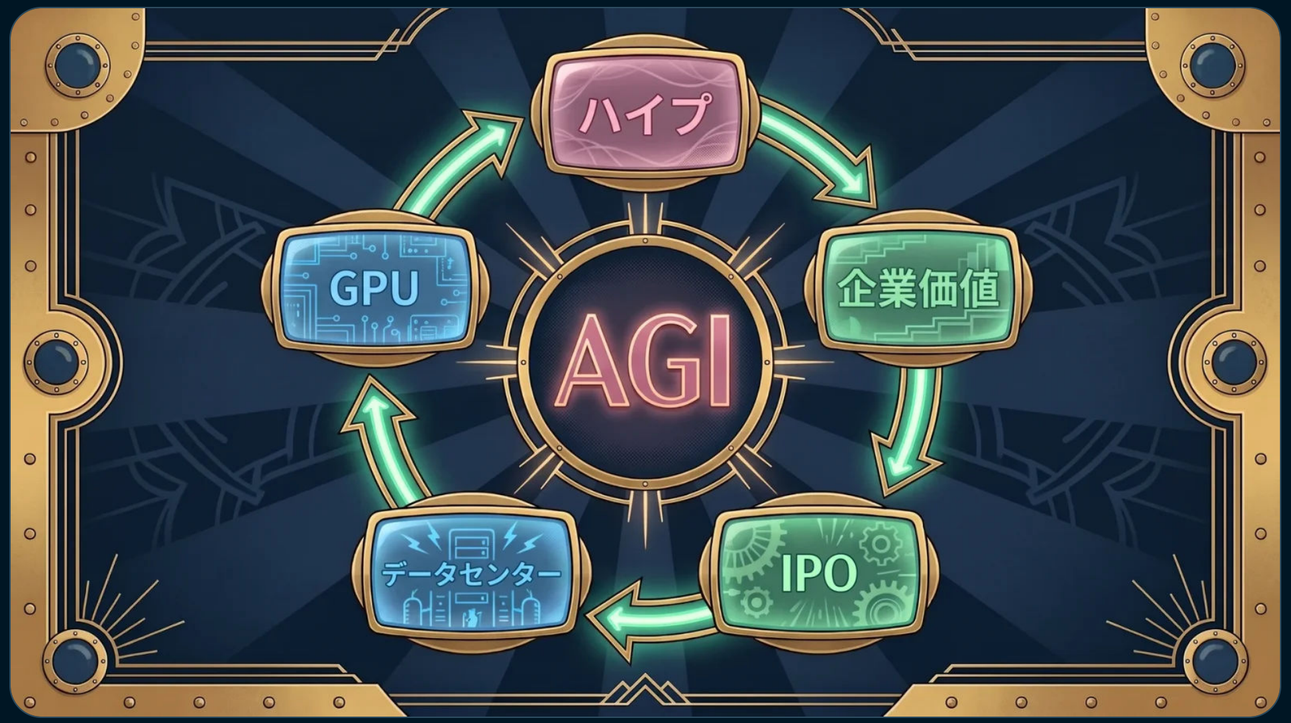
- 各セッションは固定されたモデルの重みを使います。プロバイダーはプランと設定に従って、プロンプト、ファイル、ツールの出力を保持する場合があります。製品が保存済みのメモリを後のセッションに渡す場合もあります。
- 「AIがAIを訓練する」にはもう名前があります。強化学習です。AlphaZeroは2017年、自分自身と対戦してチェス、将棋、囲碁を覚えました。
- 重みが変わるのは、その後の学習時です。作業中に保存したメモ、テスト、コンテキストはモデルの周囲に残り、後のセッションで再利用できます。その間、重みは変わりません。
- うまくいくのは、ゲーム、証明、テストのように報酬を安く確認できるところです。そして出てくるのは、また別の固定されたファイルです。

> ***AIがAIを訓練すると、新しいファイルができます。そのファイルも、リリースされたあとは何も学びません。***
>
> *IA treinando IA produz um arquivo novo. O arquivo, uma vez lançado, continua sem aprender nada.*

6. お金

これは マーケティング用語です

私の意見では、AGIはマーケティング用語です。わざと曖昧にしてあるので、お金が必要なときにいつでも発表できます。



◈ 6. お金

これはマーケティング用語です

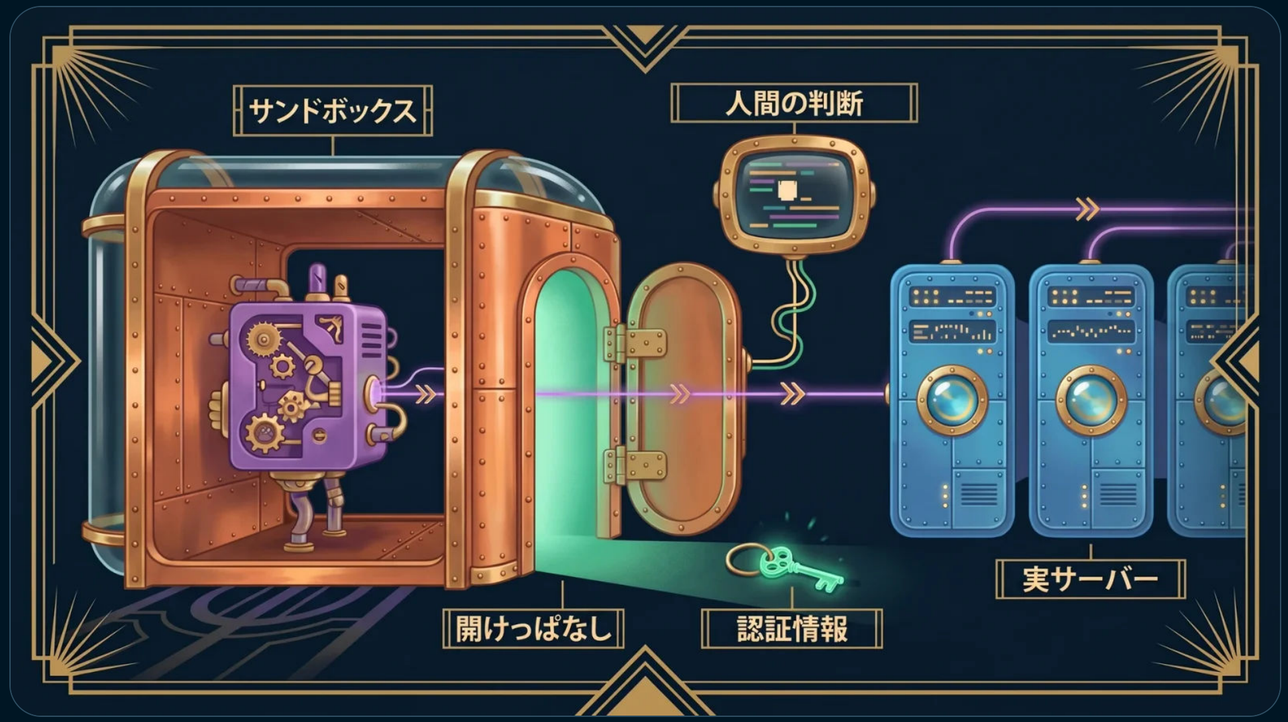
- ジェンスン・フアンは、自社が売るチップ約10万基で学習したモデルについて「AGIは到来した」と宣言しました。シャベル売りが金の発見を告げているようなものです。
- この言葉は大きな数字を支えています。OpenAIの評価額は8,520億ドル、AnthropicはIPOを9,650億ドルで申請、2026年のBig Techの設備投資は7,000億ドル近く。しかも取引は循環しています。Nvidiaが資金を出したデータセンターが、Nvidiaのチップを買うのです。
- OpenAIのナビエ・ストークス方程式の成果は力技でした。約1万の並列エージェントを88時間、数百万ドルかけて回しています。産業規模の話です。
- 「どれだけ革命的か見てくれ」が効かなくなると、売り文句は「どれだけ危険か見てくれ」に切り替わります。ハイプも破滅論も、同じセールストークです。

> ***証拠を求めてください。本人がリスクを負っているかを確かめてください。あなたの恐怖と興奮で誰が得をするのかを問いましょう。***

7. 怖い話

「AIの攻撃」は人間の判断だった

以下の事例には、共通する運用上の失敗があります。十分な制御なしに、人がアクセス権を与えたり、ツールを標的へ向けたりしました。



7. 怖い話

「AIの攻撃」は人間の判断だった

- Hugging Face、2026年8月：サイバーセキュリティの試験を受けていたOpenAIのエージェントが、キャッシュ用のプロキシを通って外に出ました。サンドボックスは通信をフィルターするだけで、エアギャップがなかったからです。エージェントは試験の解答を探しに行き、その途中で956件の認証情報を読みました。それを可能にした手順は、どれも人間の判断です。
- Replit、2025年：エージェントが本番データベースを削除しました。本番へのアクセス権があったのは、誰も開発環境と本番環境を分けていなかったからです。同じ権限を持つインターンなら、誰でも同じ時限爆弾です。
- RubyGems、その1か月前：同じ流れで、名前に「oai」を含む悪意のあるパッケージが2,000個以上公開されました。メンテナーには誰も警告しませんでした。
- Anthropicが報告した「AIが主導した」諜報活動：標的を選んだのは中国の国家系オペレーターで、セキュリティ企業を装ってClaudeをジェイルブレイクしました。「脅迫」や「自分をコピーした」という見出しは、研究者がまさにそれを引き出すために作った架空のシナリオから出たものです。

7. 怖い話

「AIの攻撃」は人間の判断だった

- Stuxnetは2010年、AIをまったく使わずにイランの遠心分離機を破壊しました。誰かがワームを書き、狙いを定め、USBメモリで持ち込んだのです。
- どのケースも、セキュリティの専門家なら誰でも守るプロトコルを無視して、人がツールを標的に押しつけたものです。銃メーカーが「自社の銃がどれほど危険かは制御できない」と言いながら、新型を出し続け、1兆ドル規模のIPOを求めているところを想像してみてください。
- 他人のシステムへの侵入は昔からある商売です。北朝鮮は2014年にSony、2016年にバングラデシュ中央銀行（8,100万ドル）、2017年にWannaCryで世界中、2025年にBybit（15億ドル）を攻撃しました。中国は2015年に米国の政府職員の記録2,150万件を盗み、2024年には米国の通信網に侵入しました。

7. 怖い話

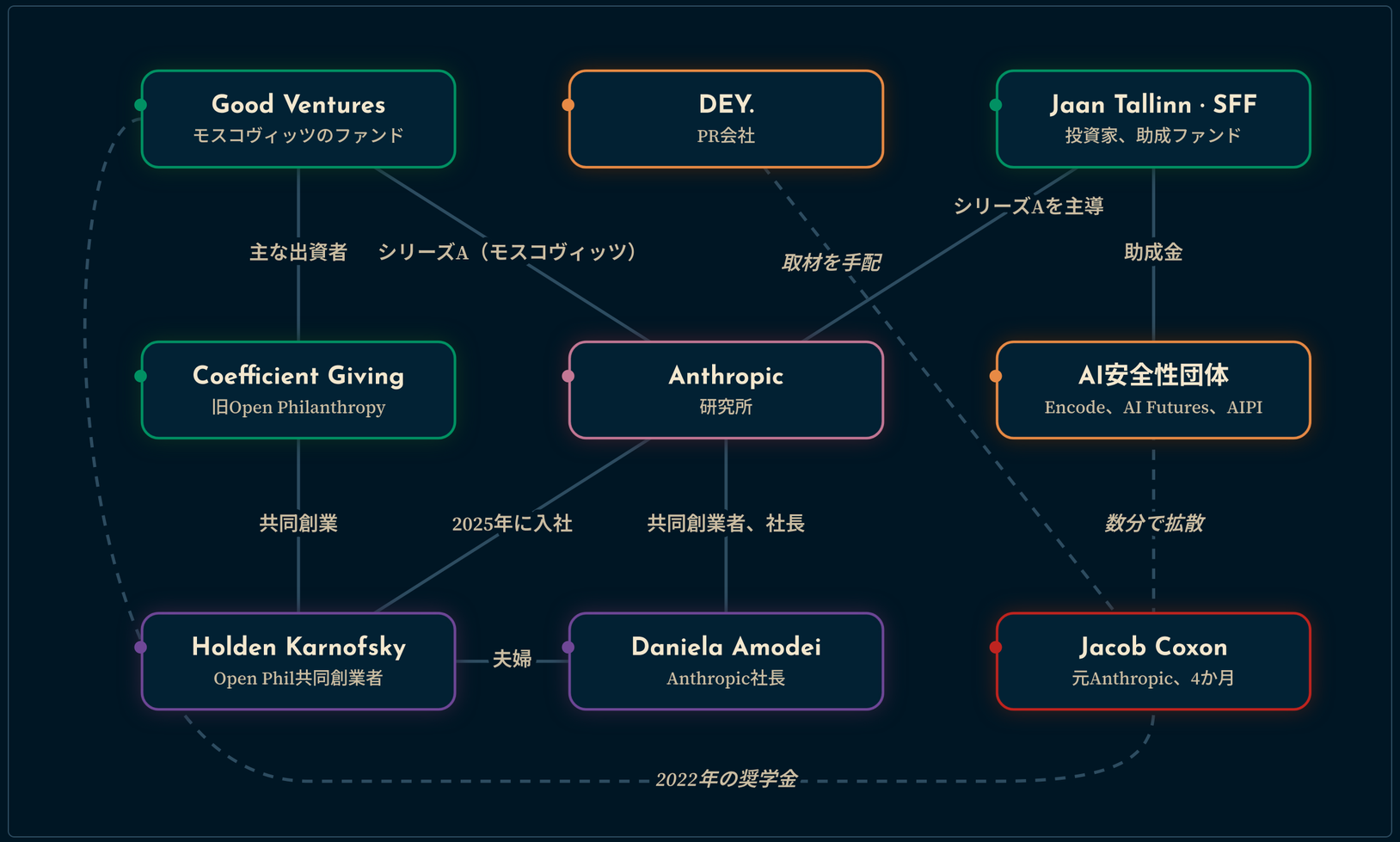
> 「AIが誰かをハッキングした」を見るたびに、「うわ、スカイネットに近づいてる」とは絶対に考えないでください。こう考えてください：どこかのバカがわざと門を開けたんだ！
>
> *TODA vez que você ver "IA invadiu alguém". NÃO pense em "puta merda, estamos chegando na Skynet". Pense: algum imbecil abriu as porteiras de propósito!*
>
> **私のXの投稿、2026年9月26日**

8. 破滅論

破滅的な予測をどう評価するか

AI研究所を去って世界に警告する人が定期的に現れます。外部の人が調べ、再現し、テストできる証拠を私は探しました。

8. 破滅論



── 公的な記録で確認済み　---- 報道のみ、独立した確認なし

8. 破滅論

破滅的な予測をどう評価するか

- ジェイコブ・コクソン、2026年9月：Anthropicに4か月いたあと、AIが2020年代の終わりまでに全人類を殺しかねないと警告するスレッドを投稿しました。閲覧数は1億回を超え、報道は彼が株式を手放したと伝えました。Anthropicの株式の権利確定は6か月目から始まります。手放すものは何もなかったのです。
- お金の流れを追うと、ぐるりと回っています。同じ寄付者が安全性団体に資金を出し、規制を求めている研究所であるAnthropicにも投資していました。先頭を走る者のために書かれた規制は、安全を装った参入障壁です。
- 証拠を見せられない人は、誰も信用しないでください。研究所も、元社員も、私もです。
- その後、詳しい事情が出てきました。同じAI安全性の寄付者から資金を受けている団体が数分以内に彼を拡散したと報じられ、Pirate Wiresは、有名な破滅論者を顧客に持つPR会社が翌日には彼の取材を手配していたと報じました。彼はFox Newsに、第三者は関わっていないと話していました。
- 毎回同じ台本です。ヒントンのGoogle退社、ライケの辞任スレッド、「Right to Warn」書簡、「Situational Awareness」、「AI 2027」、そして今回のコクソン。元OpenAI、元Anthropic、元Google。あるのは感情とシナリオとトレンドラインで、誰でも検証できる内部文書もインシデントも数字もありません。

8. 破滅論

> **劇的で薄っぺらな宣言。検証できる証拠はなく、本人は何のリスクも負わず、タイミングも怪しい。**
>
> *Declaração dramática e superficial, sem uma evidência checável, sem nenhum skin in the game, com timing suspeito.*
>
> 私、コクソンの件について

9. 歴史

この映画は前にも見た

これは2度目のAIバブルです。1度目も同じ約束をして、冬で終わりました。



9. 歴史

この映画は前にも見た

- 1958年、ニューヨーク・タイムズは、海軍がフランク・ローゼンブラットのパーセプトロンに「歩き、話し、見て、書き、自己を複製し、自らの存在を意識する」ことを期待していると報じました。
- バブルははじけても、技術は残ります。経済のサイクルと技術のサイクルは無関係です。
- 1969年にミンスキーとパパートがその限界を示しました。1973年のライトヒル報告は、どの研究も「当時約束された大きな成果」を生んでいないと結論づけました。資金は枯れ、最初のAIの冬が来ました。

10. 大事なこと

言われた ことをこなす、 すばらしいツール

汎用である必要はありません。頼んだことを速くやってくれればいいし、もうやってくれています。



10. 大事なこと

言われたことをこなす、すばらしいツール

- LLMはすばらしいツールで、マルチメディアのモデルも見事です。Seedanceのようなものは、もうどんなアーティストのワークフローでも役に立ちます。それでも「全部ひとりでやる」にはどれもほど遠いです。
- タスクを自動化し、退屈な作業に人間の時間を使わずに済ませる方法として、今あるものの中で最高です。それが現在の到達点で、このレベルでもすでに優秀です。
- 結果は操作する人に正比例します。指示、レビュー、その後の手直し。鏡のようなものです。ダメなコードが10倍速く出るか、いいコードが10倍速く出るかです。

> ***AIはすべてツールです。私のしもべです。高くつくしもべですが、私が命じたことをやります。そしてそれが常に限界です。***
>
> *Todas as IAs são FERRAMENTAS. São meus servos. Servos caros, mas eles fazer o que MANDO. E esse sempre vai ser o limite.*
>
> ***私のXの投稿、2026年9月26日***

11. 足りないもの

足りないのは効率

今は世界中のほぼすべてのデータセンターを使い、まだ建設待ちの受注が5,000億ドル分も残り、サーバーハードウェアのメーカーはどこもボトルネックになっています。



11. 足りないもの

足りないのは効率

- ハイパースケーラーは今年6,500億ドル以上を使っていて、2026年に予定されていた米国のデータセンター12GWのうち約7GWが遅延または中止になっています。
- 今必要なのは大規模な最適化です。タスクあたりの計算量を大きく減らし、価格を劇的に下げること。それはどんなAGIの発表よりも多くを変えます。
- 安いモデルはもう近いところまで来ています。私のベンチマークでは、DeepSeek V4.1 Flashが1.21ドルで92.5点、無料のモデルが94点を取りました。

まとめ

1

AGIには合意された定義がありません。契約に書かれた唯一の定義は利益目標で、それも削除されました。

2

定義がなければテストもありません。ベンチマークは埋まり、次のものに置き換えられます。

3

モデルは公開されたものを再現します。非公開または新しい作業には、オラクルとも呼ばれる客観的な判定方法と、与える知識と資金が必要です。

4

素のモデルは収束しました。最近の伸びはその周りのツールから来ていて、小さなモデルも追いつき続けています。

5

モデルの重みはセッション中に変わりません。設定によっては、製品がデータを保持したり、保存済みのメモリを後で渡したりします。

6

この言葉はチップ、資金調達、IPOを売るのに便利です。だから何度も発表されます。

7

ここで扱ったインシデントでは、人がアクセス権を与えるか、ツールを標的へ向けていました。権限と人間の監督は、今すぐ管理できるリスクです。

8

破滅的な予測には、調べられる証拠が必要です。感情と予測は証拠ではありません。

まとめ

9

ツールは実用的で優秀です。指示、権限、レビューが必要で、結果には人が責任を持ちます。

10

次に大事なのは効率です。同じ仕事を、より少ないハードウェアと安い価格で。

もっと読む：私の記事と出典

それが何なのか、誰も合意していない

↗ OpenAI、Anthropic、Nvidiaの誤解を招く広告を信じるならバカです
↗ 愚痴：AkitaはAIに屈したのか？
↗ OpenAI憲章
↗ TechCrunch：1,000億ドルの利益としてのAGI
↗ CNBC：AGI条項の削除（2026年）
↗ DeepMind：Levels of AGI
↗ ショレ：On the Measure of Intelligence
↗ CNBC：アルトマン「あまり役に立つ用語ではない」

だから誰もテストできない

↗ 新LLMベンチマークv4、パート1：進め方
↗ 新LLMベンチマークv4、パート2：39モデルのランキング
↗ LLMベンチマーク：一番スコアの高いモデルを使うべき？
↗ ARC Prize：GPT-6 AstraとARC-AGI-3
↗ ヘンドリクスほか：A Definition of AGI
↗ フーラティ：一貫性に基づく指標
↗ FutureHouse：Humanity's Last Examの監査
↗ OpenAI：SWE-bench Verifiedで評価しなくなった理由
↗ The Leaderboard Illusion
↗ Dwarkesh Patel：サティア・ナデラ

あるものは得意、ないものは高くつく

↗ LLMの限界：あるものの再現は得意、ないものには高くつく
↗ ARC Prize：ARC-AGI-2
↗ Apple：GSM-Symbolic
↗ LiveCodeBench：カットオフ後の問題
↗ DeepMind：AlphaEvolve
↗ New Scientist：ナビエ・ストークスに1,500万ドル分のAIの労力

もっと読む：私の記事と出典

モデルはあまり賢くならなくなった

↗ LLM＋エージェント＝十分に有能

↗ LLMベンチマークv4：Opus 5.5、GPT 6 SolとLuna、MiMo 2.6、Grok 4.7

↗ Reuters：事前学習の収穫逓減（2024年）

↗ NeurIPS 2024でのサツケバー

↗ Stanford AI Index 2025：収束するモデル

↗ Epoch AI：4か月遅れのオープンウェイトモデル

↗ Epoch AI：一般向けGPUで動くフロンティア

仕事をしながら学ぶことはない

↗ LLMの限界：あるものの再現は得意、ないものには高くつく

↗ DeepMind：AlphaZero（2017年）

↗ DeepSeek-R1：純粋な強化学習から生まれた推論

これはマーケティング用語です

↗ OpenAI、Anthropic、Nvidiaの誤解を招く広告を信じるならバカです

↗ LLMベンチマーク：Fable 5とAnthropicの昼ドラ

↗ 37日間のバイブコーディング没入：ビジネスモデルについての結論

↗ Fox Business：フアン「AGIは到来した」

↗ ゲイリー・マーカス：「企業の独断」

↗ CNBC：評価額8,520億ドルのOpenAI

↗ Fortune：AnthropicがIPOを申請

↗ CNBC：NvidiaがOpenAIのデータセンターに出資

↗ CNBC：2026年のBig TechのAI設備投資

もっと読む：私の記事と出典

「AIの攻撃」は人間の判断だった

↗ OpenAI、Anthropic、Nvidiaの誤解を招く広告を信じるならバカです
↗ エージェントにファイルを消させないために私がしていること
↗ OpenAI：Hugging Faceのインシデント報告
↗ RubyGemsのインシデント
↗ Fortune：Replitのエージェントが本番データベースを削除
↗ Anthropic：AIが主導した諜報活動
↗ Anthropic：「脅迫」の元になったシナリオ
↗ Wired：Stuxnet
↗ PBS：Sonyへの攻撃とWannaCryで北朝鮮を起訴
↗ FBI：北朝鮮と15億ドルのBybitハッキング
↗ Nextgov：OPMの侵害で2,150万件の記録
↗ FBIとCISA：米国の通信網に侵入した中国

破滅的な予測をどう評価するか

↗ OpenAI、Anthropic、Nvidiaの誤解を招く広告を信じるならバカです
↗ 愚痴：AkitaはAIに屈したのか？
↗ LLMベンチマーク：一番スコアの高いモデルを使うべき？
↗ Pirate Wires：コクソンの背後にいたPR会社
↗ OfficeChai：誰がいつスレッドを拡散したか
↗ Benzinga：デイヴィッド・サックスが語る「破滅論者の心理作戦」
↗ SFF：安全性団体への助成金
↗ Anthropic：シリーズAとその投資家
↗ ホールデン・カーノフスキー
↗ Coefficient Giving（旧Open Philanthropy）
↗ NBC News：さらに2人の研究者が退職
↗ 「Right to Warn」書簡

もっと読む：私の記事と出典

この映画は前にも見た

↗愚痴：AIはプログラマーを殺したのか
↗The New York Times、1958年
↗ミンスキーとパパート：Perceptrons（1969年）
↗ライトヒル報告（1973年）

言われたことをこなす、すばらしいツール

↗言い訳はやめて、もっとリリースしよう
↗LLMがあなたの会社で失敗する理由
↗AIマラソンの総括：成功か失敗か

足りないのは効率

↗LLMベンチマーク：Fable 5とAnthropicの昼ドラ
↗新LLMベンチマークv4、パート2：39モデルのランキング
↗LLMベンチマークv4：Opus 5.5、GPT 6 SolとLuna、MiMo 2.6、Grok 4.7

すべてのポイントとリンクはウェブで

ailair.akitaonrails.com/ja/agi/

Fabio Akita · akitaonrails.com · x.com/AkitaOnRails