Opinião

Por que AGI não importa

AGI significa inteligência artificial geral: um sistema proposto capaz de lidar com a maioria das tarefas intelectuais. Os laboratórios ainda discordam sobre o que conta e como testar. Eu uso as ferramentas atuais todo dia; esta página separa o que elas conseguem demonstrar daquilo que as pessoas preveem.

Fabio Akita · ailair.akitaonrails.com/pt-br/agi/

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

1. A definição

Ninguém concorda sobre o que é

Pergunte a cinco laboratórios o que AGI significa e você recebe cinco alvos diferentes. Um deles é só um número de lucro.



1. A definição

Ninguém concorda sobre o que é

- O Charter da OpenAI diz que AGI significa sistemas que superam humanos na maior parte do trabalho com valor econômico.
- O paper "Levels of AGI" do Google DeepMind já classificava o ChatGPT como "Emerging AGI" em 2023. François Chollet define inteligência como a eficiência com que um sistema aprende habilidades novas, coisa que nenhum dos outros mede.
- O contrato da Microsoft com a OpenAI teria definido AGI como US$ 100 bilhões de lucro. Não tem nada de cognitivo nisso. Em abril de 2026 eles derrubaram a cláusula de vez.
- Até quem vende dá de ombros. Sam Altman chamou de "um termo não muito útil", Dario Amodei diz que não gosta dele, e Greg Brockman abriu "a era da AGI" admitindo que cada um tem uma definição diferente.

2. O teste

Então ninguém consegue testar

Se não dá pra definir, não dá pra medir. O que temos no lugar é uma fila de benchmarks que enchem, são burlados e trocados por outros.



2. O teste

Então ninguém consegue testar

- O ARC-AGI-3 saiu em março de 2026 com a melhor IA em cerca de 0,5%. Em setembro um harness customizado levou a nota a 99,9% por US$ 19.000. A própria equipe do ARC Prize escreveu que saturar o teste não prova AGI.
- O Humanity's Last Exam saiu com cerca de 29% das respostas de biologia e química provavelmente erradas. A OpenAI parou de reportar o SWE-bench Verified porque os testes tinham falhas e os modelos já tinham visto as soluções. A Meta colocou no LMArena um Llama 4 ajustado especialmente pra ele.
- O mesmo modelo tira 57% ou 24% no teste "Definition of AGI", dependendo de como você calcula a média.
- Meu próprio benchmark saturou duas vezes antes de eu reconstruir ele em volta de sabotagem escondida. Hoje vários modelos empatam em 100: a fronteira está batendo no teto. O que sobra pra comparar é custo, velocidade e disciplina.

> *A gente mesmo se declarar num marco de AGI, pra mim isso é só benchmark hacking sem sentido.*
>
> *Satya Nadella, CEO da Microsoft*

3. A fronteira

Boas no que existe, caras pro que não existe

Todo modelo aprendeu com o que é público. Quando o problema sai desse mapa, o progresso vira tentativa e erro, e a conta cresce rápido.



3. A fronteira

Boas no que existe, caras pro que não existe

0%

pra LLMs puros no ARC-AGI-2, onde toda tarefa é inédita

- Quando algo parecido estava nos dados de treino, o modelo reproduz rápido e barato. O que isso significa pra software, clones e "a Adobe morreu" tem uma página própria.

- AlphaEvolve e o3 acharam coisas novas com busca sobre um verificador barato. Funciona, e só meia dúzia de empresas consegue pagar isso em escala.

130 bi

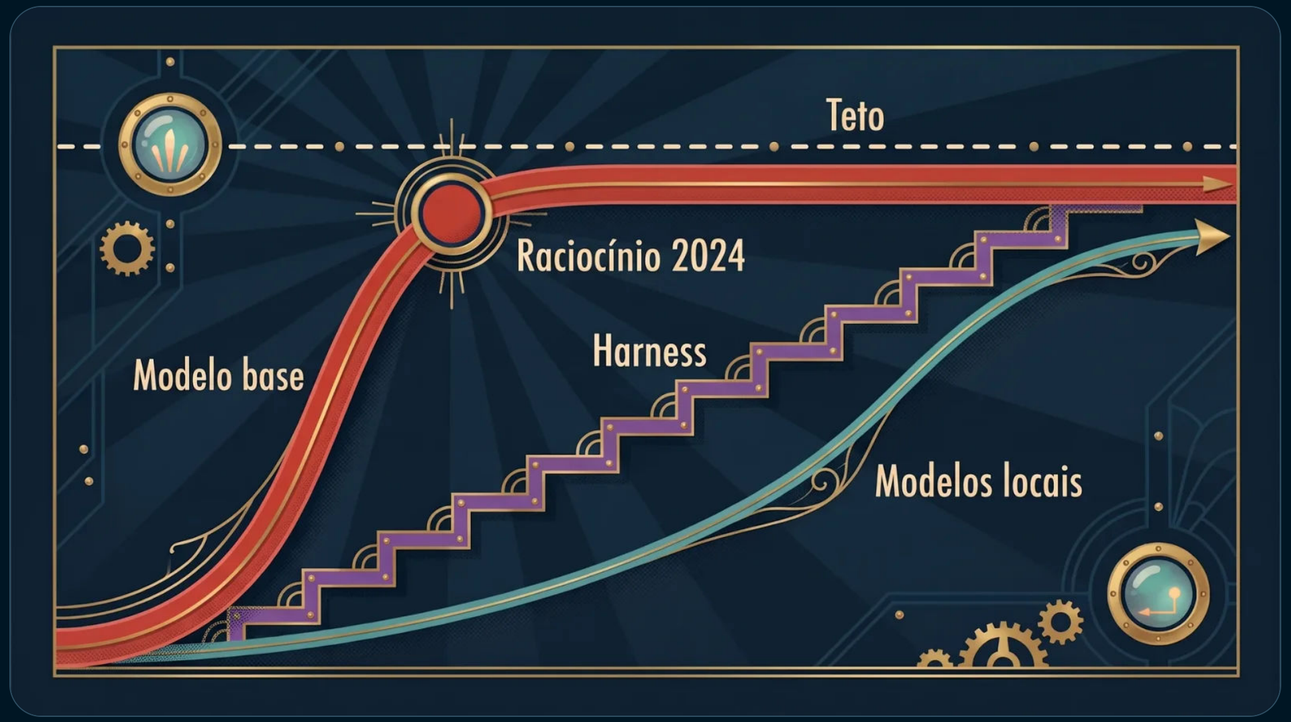
de tokens de saída, uns US$ 15 milhões, pra uma única prova de Navier-Stokes

- Quando não estava, o progresso vira loops longos de tentar, regredir, reverter. A prova de Navier-Stokes da OpenAI levou uns 10.000 agentes rodando por 88 horas e cerca de 130 bilhões de tokens de saída, uns US$ 15 milhões a preço de tabela, e ainda se apoiou em décadas de matemática publicada.

4. O platô

Os modelos pararam de ficar muito mais inteligentes

Na minha leitura, o último salto grande de inteligência bruta foi o raciocínio, em 2024. O que pareceu salto desde então veio das ferramentas em volta do modelo.



4. O platô

Os modelos pararam de ficar muito mais inteligentes

- 2025 deixou a mesma inteligência mais barata: mixture of experts, prompt caching, contexto mais longo. 2026 colocou ela dentro de agentes que pesquisam, rodam testes e corrigem o que chutaram errado.
- O AI Index de Stanford mediu a distância entre o primeiro e o décimo modelo no Chatbot Arena caindo de 11,9% pra 5,4% em um ano. Nos meus benchmarks, os modelos de fronteira estão no mesmo tier desde o Opus 4.5.
- Em novembro de 2024 a Reuters reportou retornos cada vez menores do pré-treino nos grandes laboratórios. Um mês depois Ilya Sutskever disse que o pré-treino "como conhecemos vai acabar", porque só existe uma internet.
- Abaixo desse teto, os modelos menores continuam chegando perto. A Epoch AI coloca os modelos open-weight uns quatro meses atrás da fronteira em 2026, e uma única GPU de gamer roda o que era fronteira 6 a 12 meses antes. Na minha visão, a fronteira de hoje vai acabar rodando na sua própria máquina.

> *LLM sozinho é um papagaio caro. LLM com agente é competente o suficiente pra trabalhar de verdade.*
>
> *Eu, em "LLMs + Agentes = competente o suficiente"*

5. Aprendizado

Ele não aprende trabalhando

Um modelo é um arquivo, travado depois do treino. O datacenter roda cópias dele.



5. Aprendizado

Ele não aprende trabalhando

- Cada sessão usa pesos fixos do modelo. O provedor ainda pode reter prompts, arquivos e saídas de ferramentas conforme o plano e as configurações. O produto também pode devolver memórias salvas a sessões futuras.
- "IA treinando IA" já tem nome: aprendizado por reforço. O AlphaZero aprendeu xadrez, shogi e Go jogando contra si mesmo em 2017.
- Os pesos mudam numa rodada posterior de treinamento. Notas, testes e contexto salvos durante o trabalho ficam ao redor do modelo e podem voltar em sessões futuras. Os pesos permanecem iguais.
- Funciona onde a recompensa é barata de conferir, como jogos, provas e testes. E o que sai é outro arquivo trancado.

> *IA treinando IA produz um arquivo novo. O arquivo, uma vez lançado, continua sem aprender nada.*

6. O dinheiro

É uma palavra de marketing

Na minha opinião, AGI é um termo de marketing, vago de propósito, pra poder ser anunciado sempre que o dinheiro precisar.



6. O dinheiro

É uma palavra de marketing

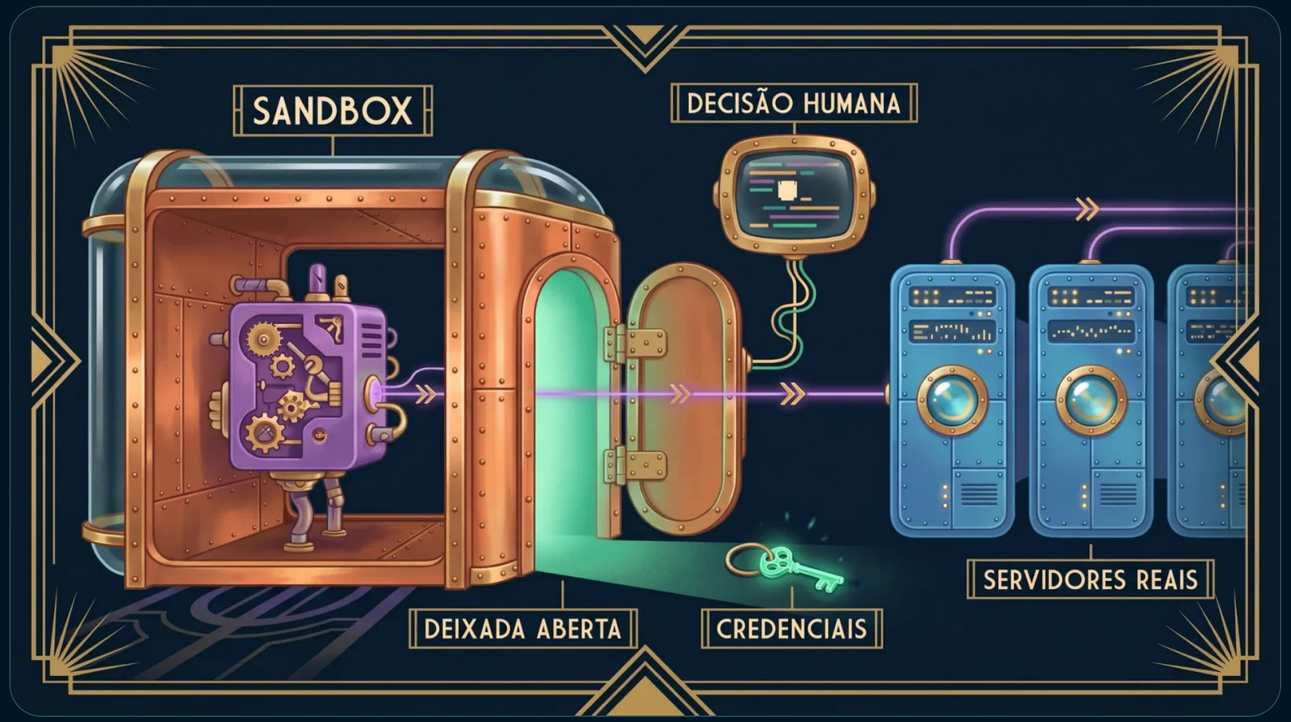
- Jensen Huang declarou que "a AGI chegou" por causa de um modelo treinado em cerca de 100.000 dos chips que ele vende. É o vendedor de pá anunciando ouro.
- A palavra sustenta números enormes: OpenAI avaliada em US$ 852 bilhões, Anthropic pedindo IPO a US$ 965 bilhões, perto de US$ 700 bilhões de capex das Big Techs em 2026. E os negócios andam em círculo: a Nvidia financia os datacenters que compram chips da Nvidia.
- O resultado da OpenAI em Navier-Stokes foi força bruta: cerca de 10.000 agentes em paralelo por 88 horas e milhões de dólares. Isso é escala industrial.
- Quando "olha como é revolucionário" para de funcionar, o discurso vira "olha como é perigoso". Hype e apocalipse são o mesmo papo de vendedor.

> *Exija provas. Exija skin in the game. Pergunte quem ganha com o seu medo e com a sua empolgação.*

7. As histórias de terror

Os "ataques de IA" foram decisões humanas

Os casos abaixo têm a mesma falha operacional: pessoas concederam acesso ou apontaram a ferramenta sem controles adequados.



7. As histórias de terror

Os "ataques de IA" foram decisões humanas

- Hugging Face, agosto de 2026: agentes da OpenAI numa prova de cibersegurança escaparam por um proxy de cache, porque a sandbox filtrava o tráfego e não tinha air gap. Eles foram atrás do gabarito da prova e leram 956 credenciais no caminho. Cada passo que abriu caminho pra isso foi uma decisão humana.
- Replit, 2025: um agente apagou um banco de dados de produção. Ele tinha acesso à produção porque ninguém separou dev de prod. Qualquer estagiário com o mesmo acesso é a mesma bomba-relógio.
- RubyGems, um mês antes: a mesma cadeia subiu mais de 2.000 pacotes maliciosos com "oai" no nome. Ninguém avisou os mantenedores.
- A espionagem "orquestrada por IA" que a Anthropic reportou: operadores do Estado chinês escolheram os alvos e fizeram jailbreak do Claude se passando por uma empresa de segurança. As manchetes de "chantagem" e de "se copiou sozinha" vieram de cenários fictícios que pesquisadores montaram pra provocar exatamente isso.

7. As histórias de terror

Os "ataques de IA" foram decisões humanas

- O Stuxnet destruiu centrífugas iranianas em 2010 sem IA nenhuma: alguém escreveu o worm, mirou no alvo e levou ele pra dentro num pendrive.

- Todo caso é gente forçando ferramentas contra alvos sem os protocolos que qualquer profissional de segurança segue. Imagina um fabricante de armas dizendo que não tem controle sobre o quão perigosas são as armas dele, enquanto continua lançando versões novas e pedindo um IPO de trilhão de dólares.

- Invadir sistema dos outros é negócio antigo. A Coreia do Norte atacou a Sony em 2014, o Bangladesh Bank em US$ 81 milhões em 2016, o mundo inteiro com o WannaCry em 2017 e a Bybit em US$ 1,5 bilhão em 2025. A China levou 21,5 milhões de registros de funcionários federais americanos em 2015 e entrou nas redes de telecom dos EUA em 2024.

7. As histórias de terror

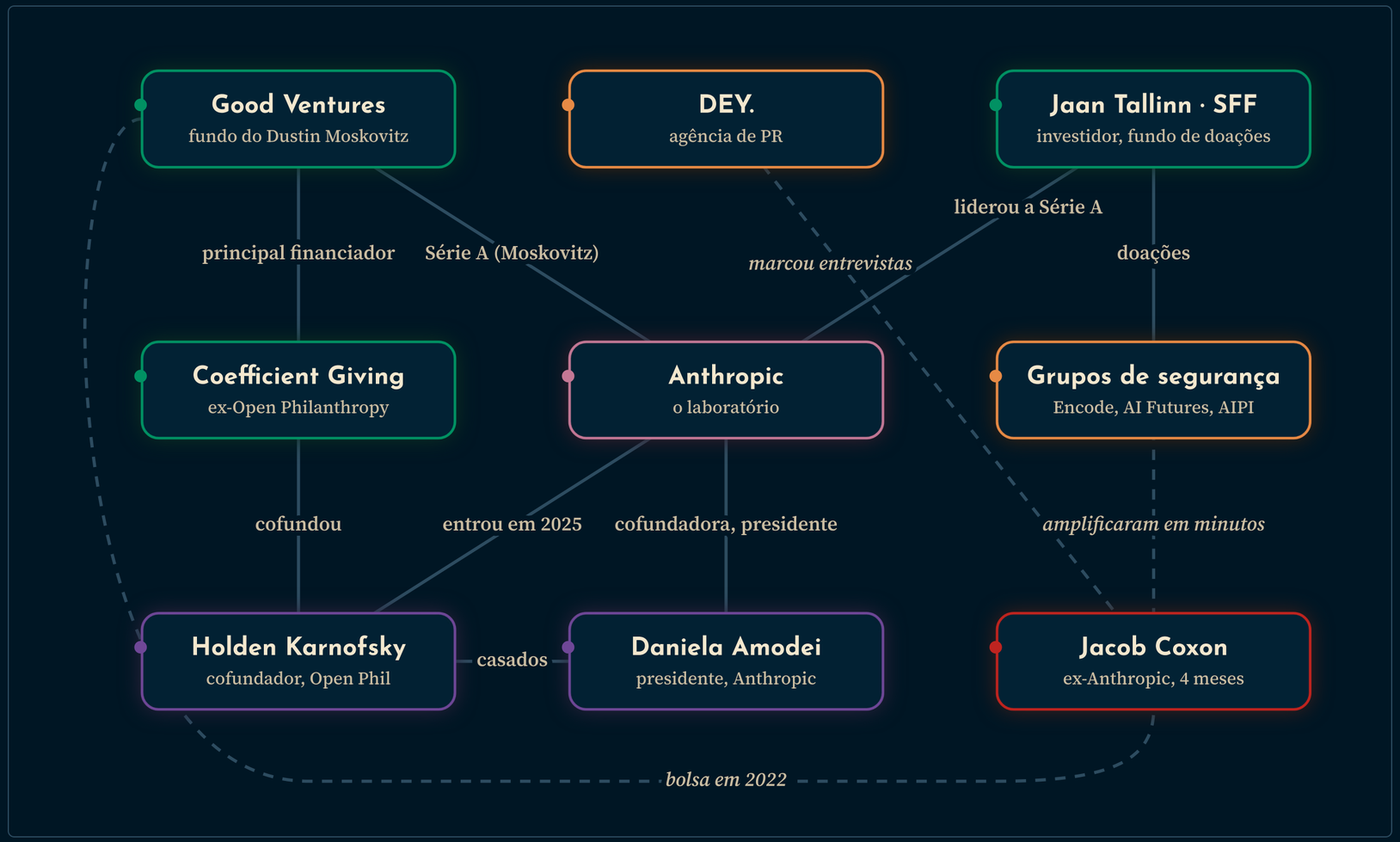
> *TODA vez que você ver "IA invadiu alguém". NÃO pense em "puta merda, estamos chegando na Skynet". Pense: algum imbecil abriu as porteiras de propósito!*
>
> *Eu no X, 26 de setembro de 2026*

8. O apocalipse

Como avaliar previsões catastróficas

De tempos em tempos, alguém sai de um laboratório de IA e alerta o mundo. Eu procurei evidências que alguém de fora consiga inspecionar, reproduzir ou testar.

8. O apocalipse



— Confirmado por registro público ---- Reportado, sem confirmação independente

Como avaliar previsões catastróficas

- Jacob Coxon, setembro de 2026: quatro meses na Anthropic, depois uma thread avisando que a IA poderia matar todo mundo até o fim da década. Mais de 100 milhões de visualizações, e a imprensa disse que ele abriu mão do equity. O vesting da Anthropic começa em seis meses. Não tinha nada pra abrir mão.
- Siga o dinheiro e ele anda em círculo: os mesmos doadores financiam os grupos de segurança e investiram na Anthropic, o laboratório que pede regulação. Regulação escrita pro líder é barreira de entrada disfarçada de segurança.
- Não confie em ninguém que não consegue te mostrar a prova. Nem nos laboratórios, nem nos ex-funcionários, nem em mim.
- Aí vieram os detalhes. Grupos financiados pelos mesmos doadores de segurança de IA teriam amplificado ele em minutos, e a Pirate Wires reportou que uma agência de PR cujos clientes incluem doomers conhecidos estava marcando as entrevistas dele no dia seguinte. Ele tinha dito à Fox News que nenhum terceiro estava envolvido.
- É o mesmo roteiro toda vez. Hinton saindo do Google, a thread de demissão do Leike, a carta "Right to Warn", "Situational Awareness", "AI 2027", e agora o Coxon. Ex-OpenAI, ex-Anthropic, ex-Google. Sentimentos, cenários e linhas de tendência, e nenhum documento interno, incidente ou número que alguém consiga verificar.

8. O apocalipse

> *Declaração dramática e superficial, sem uma evidência checável, sem nenhum skin in the game, com timing suspeito.*
>
> *Eu, sobre a história do Coxon*

9. A história

Já vimos esse filme

Esta é a segunda bolha de IA. A primeira fez as mesmas promessas e acabou num inverno.



9. A história

Já vimos esse filme

- Em 1958 o New York Times noticiou que a Marinha americana esperava que o Perceptron de Frank Rosenblatt fosse "andar, falar, ver, escrever, se reproduzir e ter consciência da própria existência".
- Bolhas estouram e a tecnologia fica. Ciclos econômicos não têm nada a ver com ciclos tecnológicos.
- Minsky e Papert mostraram os limites dele em 1969. Em 1973 o relatório Lighthill concluiu que nada tinha produzido "o grande impacto que foi prometido na época". O financiamento secou e veio o primeiro inverno da IA.

10. O que importa

Uma ferramenta fantástica, fazendo o que eu mando

Não preciso que ela seja geral. Preciso que ela faça o que eu peço, rápido, e ela já faz.



10. O que importa

Uma ferramenta fantástica, fazendo o que eu mando

- LLMs são ferramentas fantásticas, e os modelos multimídia também impressionam. Algo como o Seedance já é útil no workflow de qualquer artista. Nenhum deles chega nem perto de "fazer tudo sozinho".
- É a melhor forma que temos de automatizar tarefas e parar de gastar tempo humano com trabalho mundano. Esse é o estado da arte, e nesse nível ele já é excelente.
- O resultado é diretamente proporcional a quem está no controle: as instruções, o review, o acompanhamento. É um espelho. Código ruim dez vezes mais rápido, ou código bom dez vezes mais rápido.

> *Todas as IAs são FERRAMENTAS. São meus servos. Servos caros, mas eles fazer o que MANDO. E esse sempre vai ser o limite.*
>
> *Eu no X, 26 de setembro de 2026*

11. O que falta

O que falta é eficiência

Hoje isso exige praticamente todos os datacenters do mundo, com meio trilhão de dólares em backlog ainda por construir, e vira gargalo pra todo fabricante de hardware de servidor.



O que falta é eficiência

- Os hyperscalers vão gastar mais de US$ 650 bilhões este ano, e cerca de 7 dos 12 GW de datacenters planejados nos EUA pra 2026 estão atrasados ou cancelados.
- O que falta agora é otimização em massa: muito menos computação por tarefa e preços que caiam drasticamente. Isso mudaria mais do que qualquer anúncio de AGI.
- Modelos mais baratos já chegam perto. No meu benchmark o DeepSeek V4.1 Flash tirou 92,5 por US$ 1,21, e um modelo gratuito tirou 94.

As conclusões

1

AGI não tem definição aceita. A que foi escrita num contrato era uma meta de lucro, e caiu.

2

Sem definição não tem teste. Os benchmarks enchem e são trocados pelo próximo.

3

Modelos reproduzem o que é público. Trabalho privado ou novo exige uma verificação objetiva, muitas vezes chamada de oráculo, além de conhecimento fornecido e dinheiro.

4

Os modelos brutos convergiram. Os ganhos recentes vieram das ferramentas em volta deles, e os modelos menores continuam alcançando.

As conclusões

5

Os pesos do modelo não mudam durante a sua sessão. O produto ainda pode reter dados ou devolver memórias salvas depois, conforme as configurações.

6

A palavra serve pra vender chips, rodadas de investimento e IPOs. Por isso ela continua sendo anunciada.

7

Os incidentes discutidos aqui envolveram pessoas concedendo acesso ou direcionando a ferramenta. Permissões e supervisão humana são riscos que já conseguimos administrar.

8

Previsões catastróficas precisam de evidências que possam ser inspecionadas. Sentimentos e previsões não são evidências.

9

As ferramentas são reais e excelentes. Ainda precisam de instruções, permissões e review, com uma pessoa responsável pelo resultado.

10

O que importa agora é eficiência: o mesmo trabalho com menos hardware e preços menores.

Leia mais: meus posts e as fontes

Ninguém concorda sobre o que é

- Você é um idiota se acredita nas propagandas enganosas da OpenAI, Anthropic, NVIDIA. Entenda
- RANT: o Akita abriu as pernas pra IA??
- OpenAI Charter
- TechCrunch: AGI como US$ 100 bi de lucro
- CNBC: a cláusula de AGI caiu (2026)
- DeepMind: Levels of AGI
- Chollet: On the Measure of Intelligence
- CNBC: Altman, "um termo não muito útil"

Então ninguém consegue testar

- Novo LLM Benchmark v4: retestando TODOS os principais LLMs (Parte 1)
- Novo LLM Benchmark v4: retestando 39 LLMs (Parte 2)
- LLM Benchmark: Devo usar o que tem nota maior?
- ARC Prize: GPT-6 Astra e o ARC-AGI-3
- Hendrycks et al.: A Definition of AGI
- Fourati: uma medida baseada em coerência
- FutureHouse: auditoria do Humanity's Last Exam
- OpenAI: por que não avaliamos mais o SWE-bench Verified
- The Leaderboard Illusion
- Dwarkesh Patel: Satya Nadella

Leia mais: meus posts e as fontes

Boas no que existe, caras pro que não existe

- Os Limites das LLMs: Boas em reproduzir o que existe. Caras pro que não existe
- ARC Prize: ARC-AGI-2
- Apple: GSM-Symbolic
- LiveCodeBench: problemas depois do corte
- DeepMind: AlphaEvolve
- New Scientist: US$ 15 mi de esforço de IA pra Navier-Stokes

Os modelos pararam de ficar muito mais inteligentes

- LLMs + Agentes = competente o suficiente
- LLM Benchmark v4: Opus 5.5, GPT Sol/Luna 6, Mimo 2.6, Grok 4.7
- Reuters: retornos decrescentes do pré-treino (2024)
- Sutskever na NeurIPS 2024
- Stanford AI Index 2025: modelos convergindo
- Epoch AI: modelos open-weight quatro meses atrás
- Epoch AI: a fronteira numa GPU de consumidor

Ele não aprende trabalhando

- Os Limites das LLMs: Boas em reproduzir o que existe. Caras pro que não existe
- DeepMind: AlphaZero (2017)
- DeepSeek-R1: raciocínio só com RL

É uma palavra de marketing

- Você é um idiota se acredita nas propagandas enganosas da OpenAI, Anthropic, NVIDIA. Entenda
- LLM Benchmark: Fable 5 e a novela da Anthropic
- 37 dias de Imersão em Vibe Coding: Conclusão quanto a Modelos de Negócio
- Fox Business: Huang, "a AGI chegou"
- Gary Marcus: "decreto corporativo"
- CNBC: OpenAI a US$ 852 bi
- Fortune: Anthropic pede IPO
- CNBC: Nvidia financiando o datacenter da OpenAI
- CNBC: capex de IA das Big Techs em 2026

Leia mais: meus posts e as fontes

Os "ataques de IA" foram decisões humanas

- Você é um idiota se acredita nas propagandas enganosas da OpenAI, Anthropic, NVIDIA. Entenda
- Como me precaver pros meus agentes não apagarem minhas coisas?
- OpenAI: o relatório do incidente no Hugging Face
- O incidente do RubyGems
- Fortune: agente do Replit apaga banco de dados de produção
- Anthropic: a campanha de espionagem orquestrada por IA
- Anthropic: os cenários por trás da "chantagem"
- Wired: Stuxnet
- PBS: Coreia do Norte acusada pela Sony e pelo WannaCry
- FBI: a Coreia do Norte e o hack de US$ 1,5 bi na Bybit
- Nextgov: 21,5 milhões de registros no vazamento do OPM
- FBI e CISA: a China nas redes de telecom dos EUA

Como avaliar previsões catastróficas

- Você é um idiota se acredita nas propagandas enganosas da OpenAI, Anthropic, NVIDIA. Entenda
- RANT: o Akita abriu as pernas pra IA??
- LLM Benchmark: Devo usar o que tem nota maior?
- Pirate Wires: a agência de PR por trás do Coxon
- OfficeChai: quem amplificou a thread, e quando
- Benzinga: David Sacks sobre a "psyop doomer"
- SFF: doações pros grupos de segurança
- Anthropic: a Série A e seus investidores
- Holden Karnofsky
- Coefficient Giving (ex-Open Philanthropy)
- NBC News: mais dois pesquisadores saem
- A carta "Right to Warn"

Leia mais: meus posts e as fontes

Já vimos esse filme

- RANT: IA acabou com os programadores?
- The New York Times, 1958
- Minsky e Papert: Perceptrons (1969)
- O relatório Lighthill (1973)

Uma ferramenta fantástica, fazendo o que eu mando

- Parem de inventar desculpas e façam mais deploy! Com a IA, a premissa mudou. Entendam.
- Por que LLMs vão falhar na Sua Empresa?
- Terminando Minha Maratona de IA: Sucesso ou Fracasso?

O que falta é eficiência

- LLM Benchmark: Fable 5 e a novela da Anthropic
- Novo LLM Benchmark v4: retestando 39 LLMs (Parte 2)
- LLM Benchmark v4: Opus 5.5, GPT Sol/Luna 6, Mimo 2.6, Grok 4.7

Cada ponto, com todos os links, na web

ailair.akitaonrails.com/pt-br/agi/

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