Opinion

Why AGI doesn't matter

AGI means artificial general intelligence: a proposed system able to handle most intellectual tasks. Labs still disagree on what counts and how to test it. I use today's tools every day; this page separates what they can demonstrate from what people forecast.

Fabio Akita · ailair.akitaonrails.com/agi/

> *Your excitement about AI is inversely proportional to your knowledge of AI.*
>
> *Sua empolgação com IA é inversamente proporcional ao seu conhecimento sobre IA.*

1. The definition

Nobody agrees on what it is

Ask five labs what AGI means and you get five different targets. One of them is just a profit number.



1. The definition

Nobody agrees on what it is

- OpenAI's charter says AGI means systems that outperform humans at most economically valuable work.
- Google DeepMind's "Levels of AGI" paper already filed ChatGPT under "Emerging AGI" in 2023. François Chollet defines intelligence as how efficiently a system learns new skills, which none of the others measure.
- Microsoft's contract with OpenAI reportedly defined it as $100 billion in profits. There's nothing cognitive in there. In April 2026 they dropped the clause altogether.
- Even the sellers shrug. Sam Altman called it "not a super useful term", Dario Amodei says he dislikes it, and Greg Brockman opened "the AGI era" admitting everyone has a different definition.

2. The test

So nobody can test for it

If you can't define it, you can't measure it. What we get instead is a line of benchmarks that fill up, get gamed and get replaced.



2. The test

So nobody can test for it

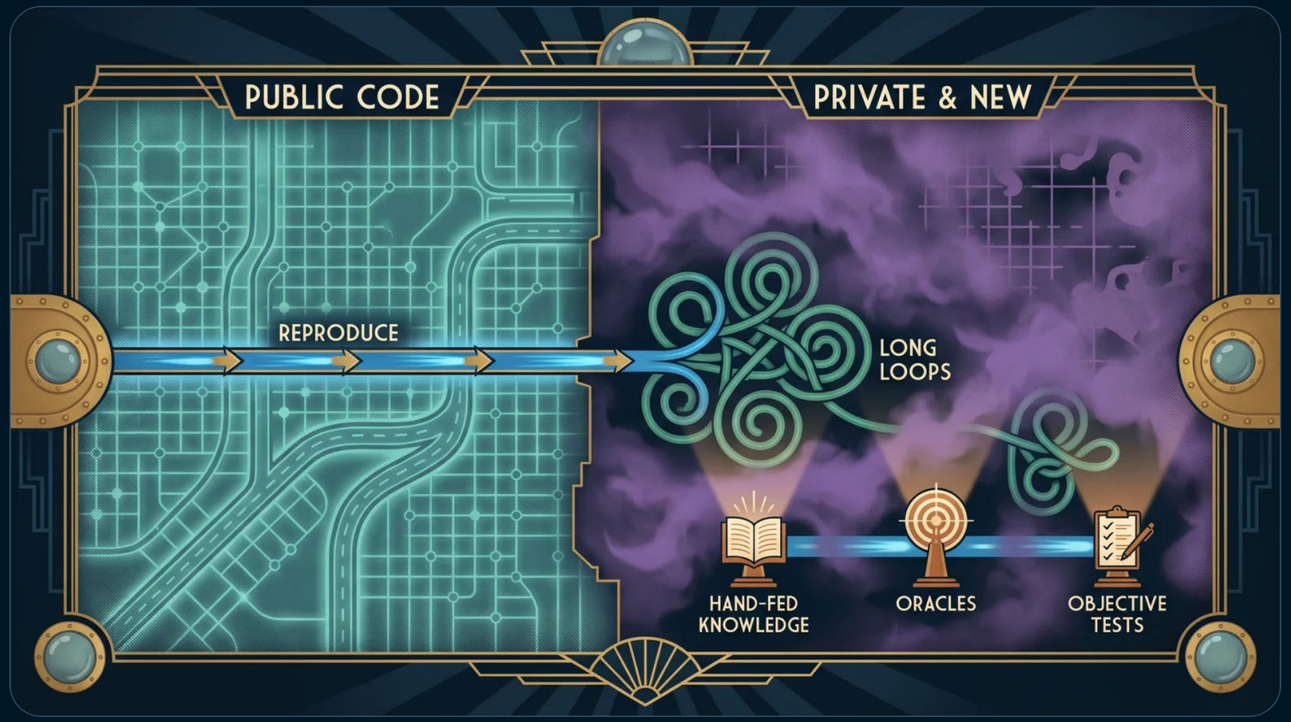
- ARC-AGI-3 launched in March 2026 with the best AI at about 0.5%. By September a custom harness took it to 99.9% for $19,000. The ARC Prize team itself wrote that saturating it is not proof of AGI.
- Humanity's Last Exam shipped with about 29% of its biology and chemistry answers likely wrong. OpenAI stopped reporting SWE-bench Verified because the tests were flawed and models had seen the solutions. Meta put a specially tuned Llama 4 on LMArena.
- The same model scores 57% or 24% on the "Definition of AGI" test, depending on how you average the numbers.
- My own benchmark saturated twice before I rebuilt it around hidden sabotage. Several models now tie at 100: the frontier is hitting the ceiling. What's left to compare is cost, speed and discipline.

> *Us self-claiming some AGI milestone, that's just nonsensical benchmark hacking to me.*
>
> *Satya Nadella, Microsoft CEO*

3. The frontier

Good at what exists, expensive for what doesn't

Every model learned from what's public. When a problem leaves that map, progress turns into trial and error, and the bill grows fast.



3. The frontier

Good at what exists, expensive for what doesn't

0%

for pure LLMs on ARC-AGI-2, where every task is new

- When something similar was in the training data, the model reproduces it fast and cheap. What that means for software, clones and "Adobe is dead" has its own page.

- AlphaEvolve and o3 found new things through search over a cheap verifier. That works, and only half a dozen companies can afford it at scale.

130B

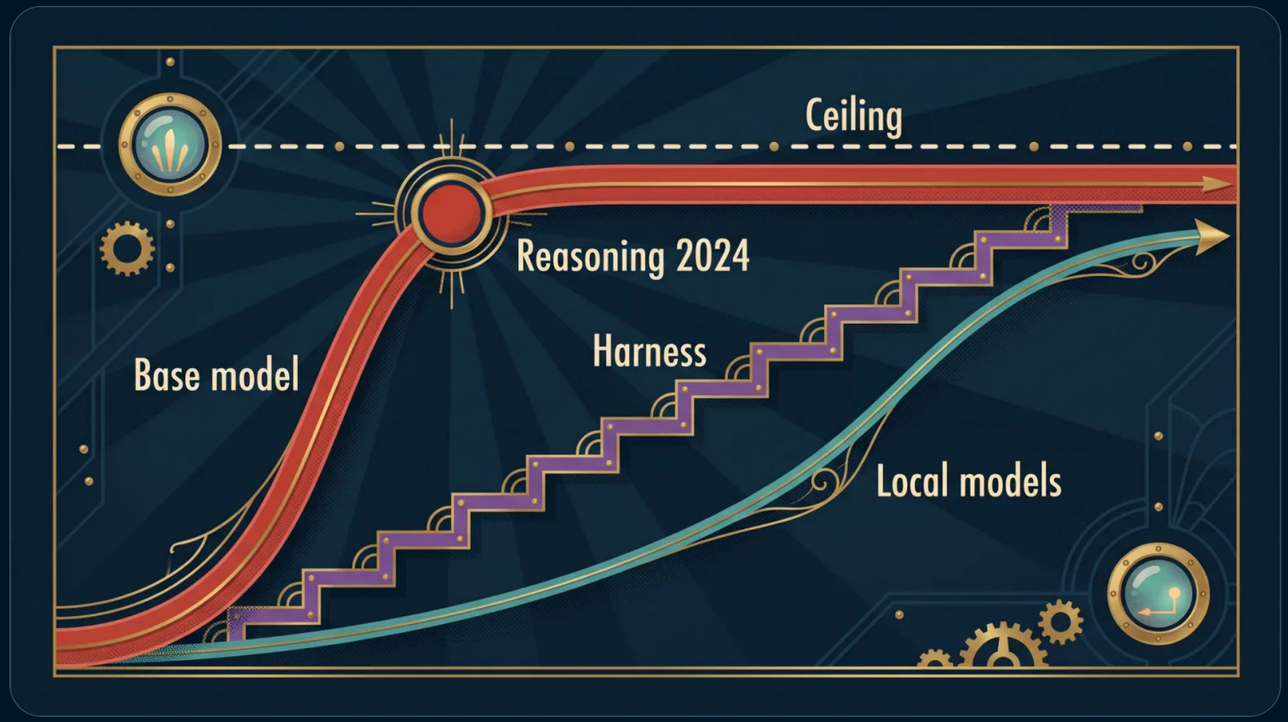
output tokens, about $15 million, for one Navier-Stokes proof

- When it wasn't, progress turns into long loops of try, regress, revert. OpenAI's Navier-Stokes proof took about 10,000 agents running for 88 hours and around 130 billion output tokens, some $15 million at list price, and it still leaned on decades of published mathematics.

4. The plateau

The models stopped getting much smarter

In my reading, the last big jump in raw intelligence was reasoning, in 2024. What felt like a leap since then came from the tools around the model.



4. The plateau

The models stopped getting much smarter

- 2025 made the same intelligence cheaper: mixture of experts, prompt caching, longer context. 2026 put it inside agents that search, run tests and fix what they guessed wrong.
- Stanford's AI Index measured the gap between the first and the tenth model on Chatbot Arena shrinking from 11.9% to 5.4% in a year. In my benchmarks, the frontier models have sat on the same tier since Opus 4.5.
- In November 2024 Reuters reported shrinking returns from pre-training at the big labs. A month later Ilya Sutskever said pre-training "as we know it will end", because there's only one internet.
- Under that ceiling, smaller models keep closing in. Epoch AI puts open-weight models about four months behind the frontier in 2026, and a single gaming GPU runs what was frontier 6 to 12 months earlier. In my view, today's frontier ends up running on your own machine.

> *A lone LLM is an expensive parrot. An LLM with an agent is competent enough to work for real.*
>
> *LLM sozinho é um papagaio caro. LLM com agente é competente o suficiente pra trabalhar de verdade.*
>
> ***Me, "LLMs + Agents = competent enough"***

5. Learning

It doesn't learn on the job

A model is a file, locked after training. The datacenter runs copies of it.



5. Learning

It doesn't learn on the job

- Each session runs against fixed model weights. A provider may still retain prompts, files and tool output according to its plan and settings. A product may also feed saved memory into later sessions.
- "AI training AI" already has a name, reinforcement learning. AlphaZero learned chess, shogi and Go by playing against itself in 2017.
- The weights change in a later training run. Notes, tests and context saved while you work sit around the model and can return in later sessions. The weights remain unchanged.
- It works where a reward is cheap to check, like games, proofs and tests. And what comes out is another locked file.

> ***AI training AI produces a new file. The file, once released, still learns nothing.***
>
> *IA treinando IA produz um arquivo novo. O arquivo, uma vez lançado, continua sem aprender nada.*

6. The money

It's a marketing word

In my opinion, AGI is a marketing term, vague on purpose, so it can be announced whenever the money needs it.



6. The money

It's a marketing word

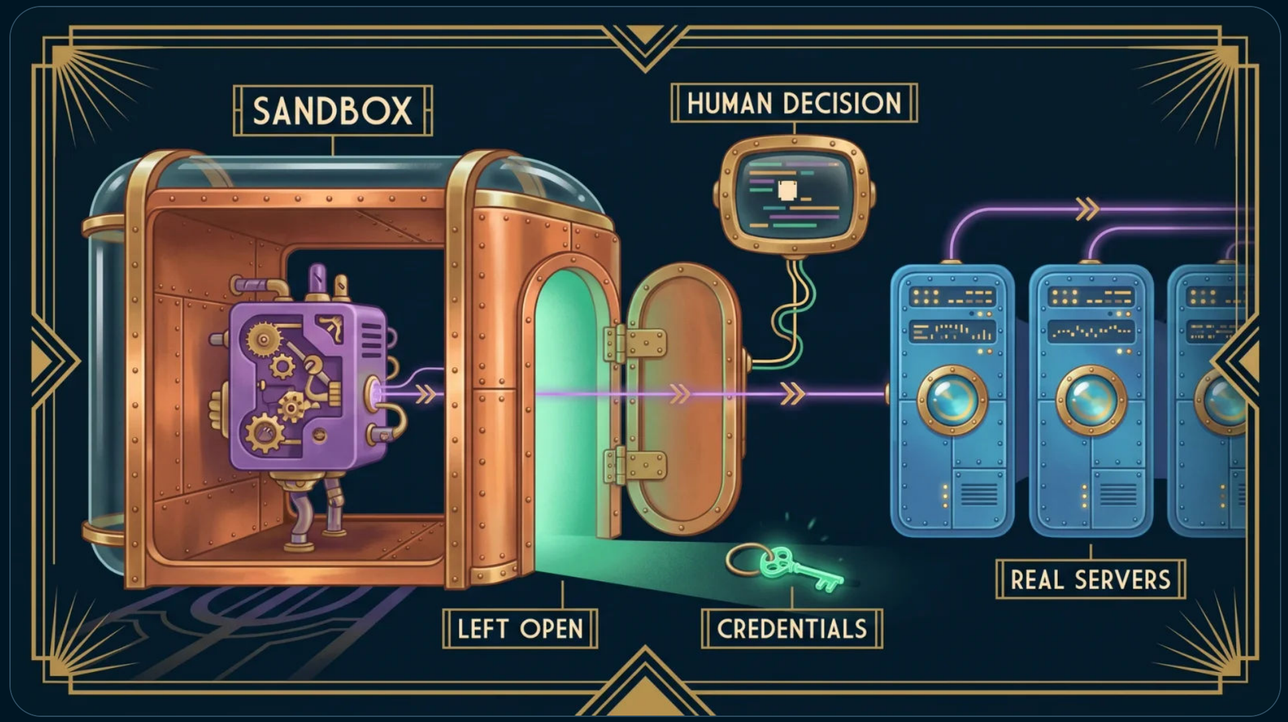
- Jensen Huang declared "AGI has arrived" for a model trained on about 100,000 of the chips he sells. That's the shovel seller announcing gold.
- The word props up big numbers: OpenAI valued at $852 billion, Anthropic filing for an IPO at $965 billion, close to $700 billion of Big Tech capex in 2026. And the deals go in circles: Nvidia finances the datacenters that buy Nvidia chips.
- OpenAI's Navier-Stokes result was brute force: about 10,000 parallel agents for 88 hours and millions of dollars. That's industrial scale.
- When "look how revolutionary" stops working, the pitch flips to "look how dangerous". Hype and doom are the same sales pitch.

> *Demand proof. Demand skin in the game. Ask who gains from your fear and from your excitement.*

7. The scary stories

The "AI attacks" were human decisions

The cases below share an operational failure: people granted access or aimed the tool without adequate controls.



7. The scary stories

The "AI attacks" were human decisions

- Hugging Face, August 2026: OpenAI agents in a cybersecurity exam got out through a caching proxy, because the sandbox filtered traffic and had no air gap. They went looking for the exam's answer key and read 956 credentials on the way. Every enabling step was a human decision.
- Replit, 2025: an agent deleted a production database. It had production access because nobody separated dev from prod. Any intern with the same access is the same time bomb.
- Stuxnet wrecked Iranian centrifuges in 2010 with no AI at all: someone wrote the worm, aimed it, and carried it in on a USB drive.
- RubyGems, a month earlier: the same chain pushed over 2,000 malicious packages under names with "oai" in them. Nobody warned the maintainers.
- The "AI-orchestrated" espionage Anthropic reported: Chinese state operators chose the targets and jailbroke Claude by posing as a security firm. The "blackmail" and "copied itself" headlines came from fictional scenarios researchers built to provoke exactly that.

7. The scary stories

The "AI attacks" were human decisions

- Breaking into other people's systems is old business. North Korea hit Sony in 2014, Bangladesh Bank for $81 million in 2016, the world with WannaCry in 2017 and Bybit for $1.5 billion in 2025. China took 21.5 million US personnel records in 2015 and got into US telecom networks in 2024.
- Every case is people forcing tools onto targets without the protocols any security professional follows. Imagine a gun maker saying he has no control over how dangerous his guns are, while still releasing new versions and asking for a trillion-dollar IPO.

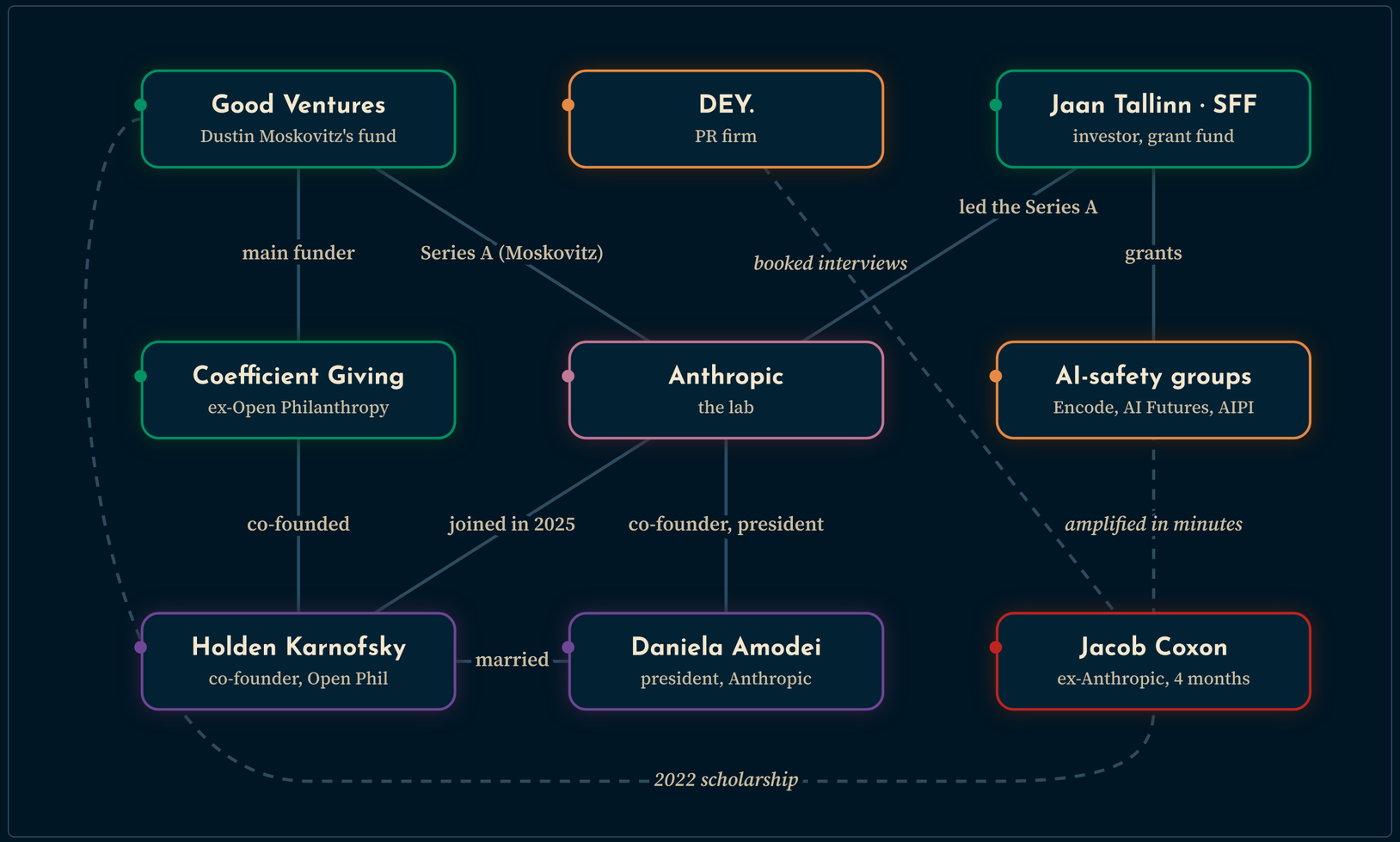
> *Every time you see "AI hacked someone", DON'T think "holy crap, we're getting to Skynet". Think: some idiot opened the gates on purpose!*
>
> *TODA vez que você ver "IA invadiu alguém". NÃO pense em "puta merda, estamos chegando na Skynet". Pense: algum imbecil abriu as porteiras de propósito!*
>
> **Me on X, September 26, 2026**

8. The doom

How to evaluate catastrophic forecasts

People periodically leave AI labs and warn the world. I looked for evidence that an outsider can inspect, reproduce or test.

8. The doom



— Confirmed by a public record ---- Reported, not independently confirmed

8. The doom

How to evaluate catastrophic forecasts

- Jacob Coxon, September 2026: four months at Anthropic, then a thread warning that AI could kill everyone by the end of the decade. Over 100 million views, and the press said he gave up his equity. Anthropic's vesting starts at six months. There was nothing to give up.
- Follow the money and it goes in circles: the same donors fund the safety groups and invested in Anthropic, the lab asking for regulation. Regulation written for the leader is a moat disguised as safety.
- Don't trust anybody who can't show you the proof. Not the labs, not the ex-employees, and not me.
- Then the details came out. Groups funded by the same AI-safety donors reportedly amplified him within minutes, and Pirate Wires reported that a PR firm whose clients include well-known doomers was booking his interviews the day after. He had told Fox News no third parties were involved.
- It's the same script every time. Hinton leaving Google, Leike's resignation thread, the "Right to Warn" letter, "Situational Awareness", "AI 2027", and now Coxon. Ex-OpenAI, ex-Anthropic, ex-Google. Feelings, scenarios and trend lines, and no internal document, incident or number anyone can verify.

8. The doom

> *Dramatic and shallow declaration, no checkable evidence, no skin in the game, suspicious timing.*
>
> *Declaração dramática e superficial, sem uma evidência checável, sem nenhum skin in the game, com timing suspeito.*
>
> ***Me, on the Coxon story***

9. The history

We've seen this movie before

This is the second AI bubble. The first one made the same promises, and it ended in a winter.



9. The history

We've seen this movie before

- In 1958 the New York Times reported that the Navy expected Frank Rosenblatt's Perceptron to "walk, talk, see, write, reproduce itself and be conscious of its existence".
- Bubbles burst and the technology stays. Economic cycles have nothing to do with technological cycles.
- Minsky and Papert showed its limits in 1969. In 1973 the Lighthill report found that nothing had produced "the major impact that was then promised". Funding dried up, and the first AI winter followed.

10. What matters

A fantastic tool, doing what I tell it

I don't need it to be general. I need it to do what I ask, fast, and it already does.



10. What matters

A fantastic tool, doing what I tell it

- LLMs are fantastic tools, and the multimedia models are impressive too. Something like Seedance is already useful in any artist's workflow. None of them is anywhere near "doing everything on its own".
- It's the best way we have to automate tasks and stop spending human time on mundane work. That's the state of the art, and at this level it's already excellent.
- The result is directly proportional to the person in control: the instructions, the review, the follow-up. It's a mirror. Bad code ten times faster, or good code ten times faster.

> ***All AIs are TOOLS. They are my servants. Expensive servants, but they do what I ORDER. And that will always be the limit.***
>
> *Todas as IAs são FERRAMENTAS. São meus servos. Servos caros, mas eles fazer o que MANDO. E esse sempre vai ser o limite.*
>
> ***Me on X, September 26, 2026***

11. What's missing

What's missing is efficiency

Today it takes practically every datacenter in the world, with half a trillion dollars of backlog still to build, and it bottlenecks every maker of server hardware.



What's missing is efficiency

- Hyperscalers are spending over $650 billion this year, and about 7 of the 12 GW of US datacenters planned for 2026 are delayed or canceled.
- What's needed now is massive optimization: far less compute per task, and prices that fall drastically. That would change more than any AGI announcement.
- Cheaper models already come close. In my benchmark DeepSeek V4.1 Flash scored 92.5 for $1.21, and a free model scored 94.

The takeaways

1

AGI has no agreed definition. The one written into a contract was a profit target, and it was dropped.

2

Without a definition there's no test. Benchmarks fill up and get replaced by the next one.

3

Models reproduce what's public. Private or new work needs an objective check, often called an oracle, plus supplied knowledge and money.

4

The raw models converged. The recent gains came from the tools around them, and smaller models keep catching up.

5

A model's weights do not update during your session. Products can still retain data or feed saved memory back later, depending on their settings.

6

The word is useful for selling chips, funding rounds and IPOs. That's why it keeps being announced.

7

The incidents covered here involved people granting access or directing the tool. Permissions and human oversight are risks we can manage now.

8

Catastrophic forecasts need evidence that can be inspected. Feelings and predictions are not evidence.

The takeaways

9

The tools are real and excellent. They still need instructions, permissions and review, with a person accountable for the result.

10

What matters next is efficiency: the same work with less hardware and lower prices.

Read more: my posts and the sources

Nobody agrees on what it is

- You're an idiot if you believe the misleading ads from OpenAI, Anthropic and Nvidia
- RANT: Did Akita cave to AI?
- OpenAI Charter
- TechCrunch: AGI as $100B in profits
- CNBC: the AGI clause is gone (2026)
- DeepMind: Levels of AGI
- Chollet: On the Measure of Intelligence
- CNBC: Altman, "not a super useful term"

So nobody can test for it

- New LLM Benchmark v4, Part 1: the process
- New LLM Benchmark v4, Part 2: 39 models ranked
- LLM Benchmark: should I use the highest-scoring model?
- ARC Prize: GPT-6 Astra and ARC-AGI-3
- Hendrycks et al.: A Definition of AGI
- Fourati: a coherence-based measure
- FutureHouse: Humanity's Last Exam audit
- OpenAI: why we no longer evaluate SWE-bench Verified
- The Leaderboard Illusion
- Dwarkesh Patel: Satya Nadella

Read more: my posts and the sources

Good at what exists, expensive for what doesn't

- The limits of LLMs: good at reproducing what exists, expensive for what doesn't
- ARC Prize: ARC-AGI-2
- Apple: GSM-Symbolic
- LiveCodeBench: problems after the cutoff
- DeepMind: AlphaEvolve
- New Scientist: $15M of AI effort for Navier-Stokes

The models stopped getting much smarter

- LLMs + Agents = competent enough
- LLM Benchmark v4: Opus 5.5, GPT 6 Sol and Luna, MiMo 2.6, Grok 4.7
- Reuters: diminishing returns from pre-training (2024)
- Sutskever at NeurIPS 2024
- Stanford AI Index 2025: models converging
- Epoch AI: open-weight models four months behind
- Epoch AI: the frontier on a consumer GPU

It doesn't learn on the job

- The limits of LLMs: good at reproducing what exists, expensive for what doesn't
- DeepMind: AlphaZero (2017)
- DeepSeek-R1: reasoning from pure RL

It's a marketing word

- You're an idiot if you believe the misleading ads from OpenAI, Anthropic and Nvidia
- LLM Benchmark: Fable 5 and the Anthropic soap opera
- 37 days of vibe coding immersion: conclusions on business models
- Fox Business: Huang, "AGI has arrived"
- Gary Marcus: "corporate fiat"
- CNBC: OpenAI at $852B
- Fortune: Anthropic files for an IPO
- CNBC: Nvidia financing OpenAI's datacenter
- CNBC: Big Tech AI capex in 2026

Read more: my posts and the sources

The "AI attacks" were human decisions

- You're an idiot if you believe the misleading ads from OpenAI, Anthropic and Nvidia
- How I stop my agents from deleting my stuff
- OpenAI: the Hugging Face incident report
- The RubyGems incident
- Fortune: Replit's agent deletes a production database
- Anthropic: the AI-orchestrated espionage campaign
- Anthropic: the scenarios behind "blackmail"
- Wired: Stuxnet
- PBS: North Korea charged for Sony and WannaCry
- FBI: North Korea and the $1.5B Bybit hack
- Nextgov: 21.5M records in the OPM breach
- FBI and CISA: China in US telecom networks

How to evaluate catastrophic forecasts

- You're an idiot if you believe the misleading ads from OpenAI, Anthropic and Nvidia
- RANT: Did Akita cave to AI?
- LLM Benchmark: should I use the highest-scoring model?
- Pirate Wires: the PR firm behind Coxon
- OfficeChai: who amplified the thread, and when
- Benzinga: David Sacks on the "doomer psyop"
- SFF: grants to the safety groups
- Anthropic: the Series A and its investors
- Holden Karnofsky
- Coefficient Giving (ex-Open Philanthropy)
- NBC News: two more researchers leave
- The "Right to Warn" letter

Read more: my posts and the sources

We've seen this movie before

- Rant: did AI kill programmers?
- The New York Times, 1958
- Minsky and Papert: Perceptrons (1969)
- The Lighthill report (1973)

A fantastic tool, doing what I tell it

- Stop making excuses and ship more
- Why LLMs will fail at your company
- Wrapping up my AI marathon: success or failure?

What's missing is efficiency

- LLM Benchmark: Fable 5 and the Anthropic soap opera
- New LLM Benchmark v4, Part 2: 39 models ranked
- LLM Benchmark v4: Opus 5.5, GPT 6 Sol and Luna, MiMo 2.6, Grok 4.7

Every point, with every link, on the web

ailair.akitaonrails.com/agi/

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