Opinión

Por qué la AGI no importa

AGI significa inteligencia artificial general: un sistema propuesto capaz de abordar la mayoría de las tareas intelectuales. Los laboratorios aún discrepan sobre qué cuenta y cómo probarlo. Uso las herramientas actuales todos los días; esta página separa lo que pueden demostrar de lo que la gente pronostica.

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

> *Tu entusiasmo por la IA es inversamente proporcional a lo que sabes de IA.*
>
> *Sua empolgação com IA é inversamente proporcional ao seu conhecimento sobre IA.*

1. La definición

Nadie se pone de acuerdo en qué es

Pregunta a cinco laboratorios qué significa AGI y tendrás cinco objetivos distintos. Uno de ellos es solo una cifra de ganancias.



1. La definición

Nadie se pone de acuerdo en qué es

- La carta fundacional de OpenAI dice que la AGI son sistemas que superan a los humanos en la mayor parte del trabajo con valor económico.
- El paper "Levels of AGI" de Google DeepMind ya clasificaba a ChatGPT como "AGI emergente" en 2023. François Chollet define la inteligencia como la eficiencia con la que un sistema aprende habilidades nuevas, algo que ninguno de los demás mide.
- Según se informó, el contrato de Microsoft con OpenAI la definía como 100.000 millones de dólares en ganancias. Ahí no hay nada cognitivo. En abril de 2026 eliminaron la cláusula por completo.
- Hasta los que la venden se encogen de hombros. Sam Altman dijo que "no es un término muy útil", Dario Amodei dice que no le gusta y Greg Brockman inauguró "la era de la AGI" admitiendo que cada uno tiene una definición distinta.

2. La prueba

Así que nadie puede medirla

Si no la puedes definir, no la puedes medir. Lo que tenemos en su lugar es una sucesión de benchmarks que se saturan, se manipulan y se reemplazan.



2. La prueba

Así que nadie puede medirla

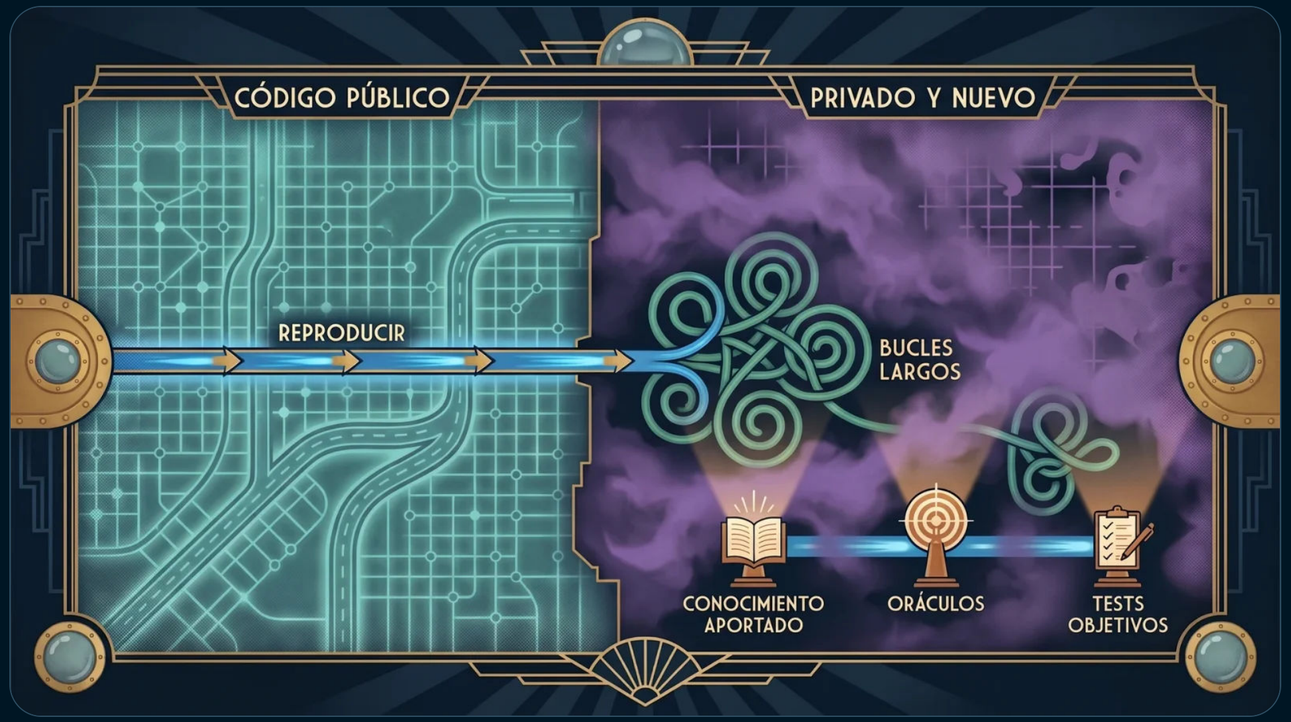
- ARC-AGI-3 salió en marzo de 2026 y la mejor IA sacaba cerca del 0,5 %. En septiembre un harness a medida la llevó al 99,9 % por 19.000 dólares. El propio equipo del ARC Prize escribió que saturarlo no prueba la AGI.
- Humanity's Last Exam salió con cerca del 29 % de sus respuestas de biología y química probablemente incorrectas. OpenAI dejó de reportar SWE-bench Verified porque los tests tenían fallas y los modelos ya habían visto las soluciones. Meta puso en LMArena un Llama 4 ajustado especialmente.
- El mismo modelo saca 57 % o 24 % en la prueba "Definition of AGI", según cómo promedies los números.
- Mi propio benchmark se saturó dos veces antes de que lo rehiciera en torno a un sabotaje oculto. Hoy varios modelos empatan en 100: la frontera está llegando al techo. Lo que queda por comparar es costo, velocidad y disciplina.

> *Que nosotros mismos proclamemos algún hito de AGI me parece un hackeo de benchmarks sin sentido.*
>
> *Satya Nadella, CEO de Microsoft*

3. La frontera

Bueno en lo que existe, caro en lo que no

Todo modelo aprendió de lo que es público. Cuando un problema sale de ese mapa, el progreso se vuelve prueba y error, y la cuenta crece rápido.



3. La frontera

Bueno en lo que existe, caro en lo que no

0 %

para LLMs puros en ARC-AGI-2, donde cada tarea es nueva

- Cuando había algo parecido en los datos de entrenamiento, el modelo lo reproduce rápido y barato. Lo que eso significa para el software, los clones y "Adobe murió" tiene su propia página.

130 mil M

tokens de salida, unos 15 millones de dólares, para una sola demostración de Navier-Stokes

- Cuando no lo había, el progreso se vuelve bucles largos de intentar, retroceder, revertir. La demostración de Navier-Stokes de OpenAI necesitó unos 10.000 agentes funcionando durante 88 horas y alrededor de 130.000 millones de tokens de salida, unos 15 millones de dólares a precio de lista, y aun así se apoyó en décadas de matemáticas publicadas.

3. La frontera

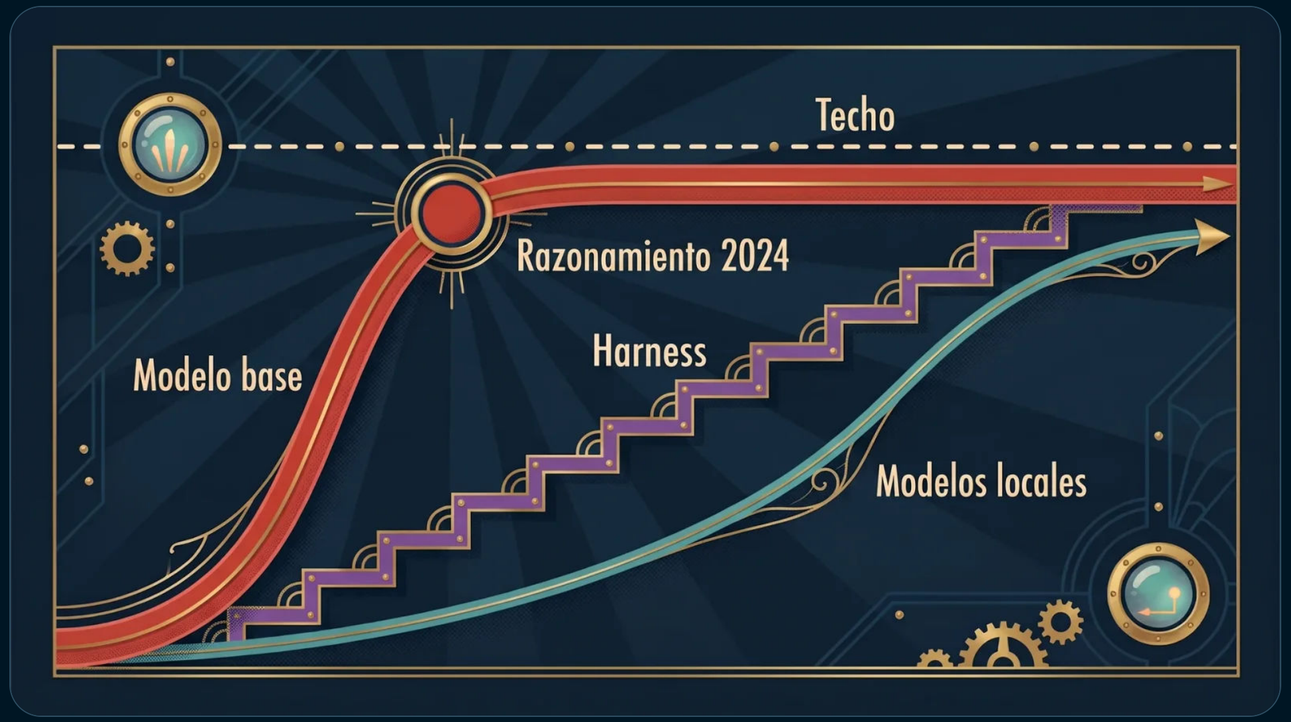
Bueno en lo que existe, caro en lo que no

- AlphaEvolve y o3 encontraron cosas nuevas con búsqueda sobre un verificador barato. Funciona, y solo media docena de empresas pueden pagarlo a escala.

4. La meseta

Los modelos dejaron de volverse mucho más inteligentes

En mi lectura, el último gran salto de inteligencia bruta fue el razonamiento, en 2024. Lo que desde entonces pareció un salto vino de las herramientas alrededor del modelo.



4. La meseta

Los modelos dejaron de volverse mucho más inteligentes

- 2025 abarató la misma inteligencia: mixture of experts, prompt caching, contexto más largo. 2026 la puso dentro de agentes que buscan, corren tests y corrigen lo que adivinaron mal.

- El AI Index de Stanford midió que la distancia entre el primer y el décimo modelo en Chatbot Arena se redujo del 11,9 % al 5,4 % en un año. En mis benchmarks, los modelos de frontera están en el mismo nivel desde Opus 4.5.

- En noviembre de 2024, Reuters informó que el preentrenamiento rendía cada vez menos en los grandes laboratorios. Un mes después, Ilya Sutskever dijo que el preentrenamiento "como lo conocemos va a terminar", porque solo hay un internet.

- Bajo ese techo, los modelos más pequeños siguen acercándose. Epoch AI pone a los modelos de pesos abiertos unos cuatro meses detrás de la frontera en 2026, y una sola GPU gamer corre lo que era frontera entre 6 y 12 meses antes. En mi opinión, la frontera de hoy terminará corriendo en tu propia máquina.

4. La meseta

> *Un LLM solo es un loro caro. Un LLM con un agente es lo bastante competente para trabajar de verdad.*
>
> *LLM sozinho é um papagaio caro. LLM com agente é competente o suficiente pra trabalhar de verdade.*
>
> ***Yo, "LLMs + agentes = lo bastante competentes"***

5. Aprender

No aprende trabajando

Un modelo es un archivo, cerrado después del entrenamiento. El datacenter corre copias de él.



5. Aprender

No aprende trabajando

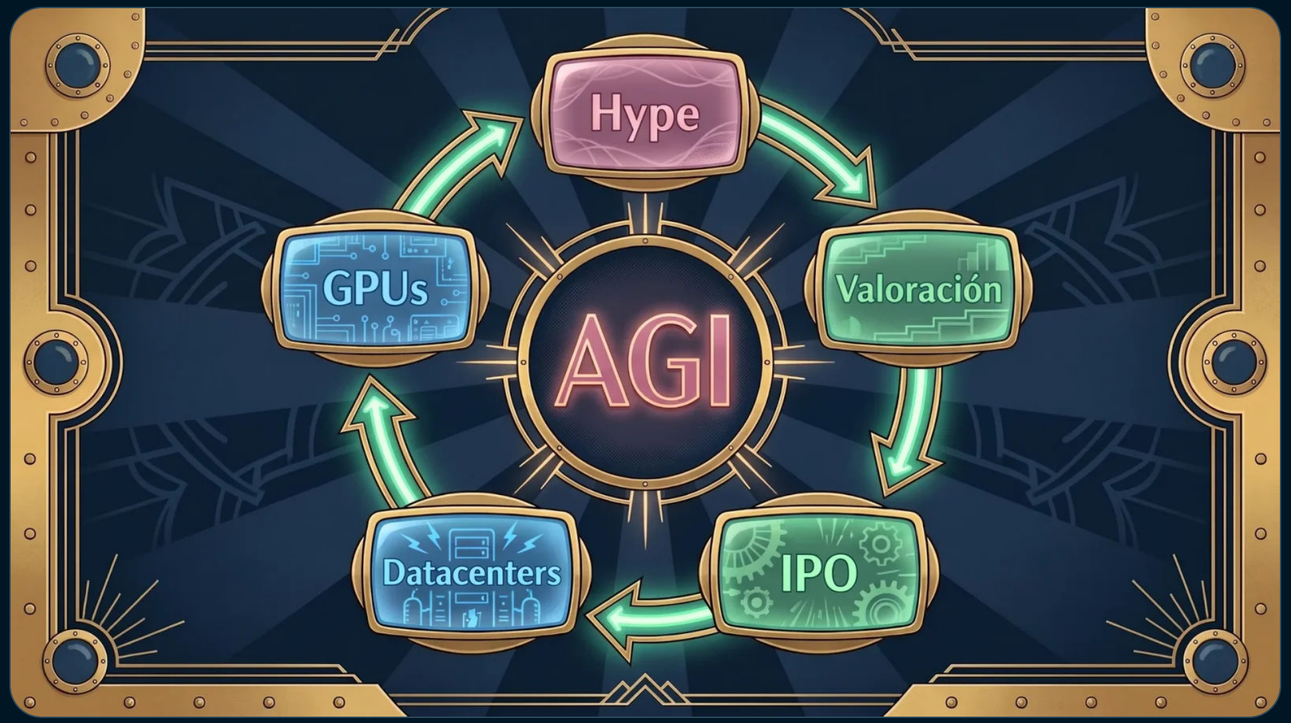
- Cada sesión usa pesos fijos del modelo. El proveedor aún puede retener prompts, archivos y resultados de herramientas según el plan y la configuración. El producto también puede devolver memoria guardada a sesiones posteriores.
- "IA entrenando IA" ya tiene nombre: aprendizaje por refuerzo. AlphaZero aprendió ajedrez, shogi y go jugando contra sí mismo en 2017.
- Los pesos cambian en un entrenamiento posterior. Las notas, los tests y el contexto guardados durante el trabajo quedan alrededor del modelo y pueden volver en sesiones posteriores. Los pesos siguen iguales.
- Funciona donde una recompensa es barata de verificar, como juegos, demostraciones y tests. Y lo que sale es otro archivo cerrado.

> ***IA entrenando IA produce un archivo nuevo. Ese archivo, una vez lanzado, sigue sin aprender nada.***
>
> *IA treinando IA produz um arquivo novo. O arquivo, uma vez lançado, continua sem aprender nada.*

6. El dinero

Es una palabra de marketing

En mi opinión, AGI es un término de marketing, vago a propósito, para poder anunciarlo cuando el dinero lo necesite.



6. El dinero

Es una palabra de marketing

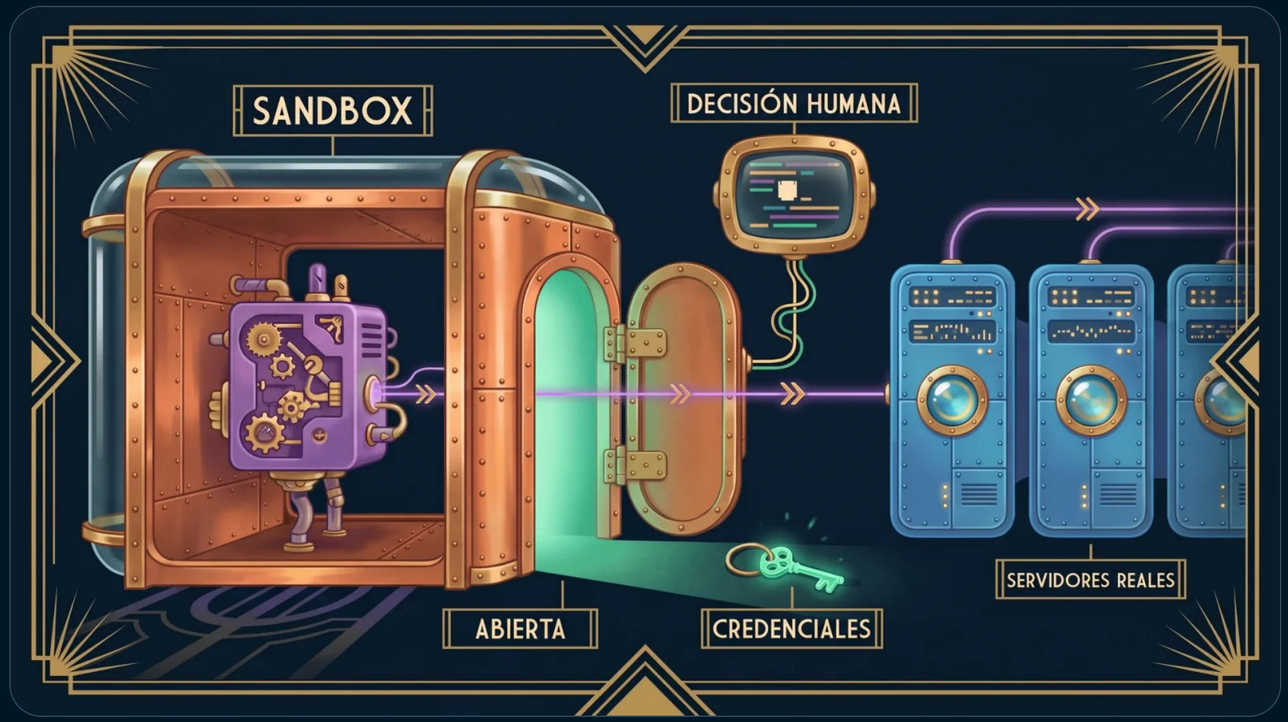
- Jensen Huang declaró que "la AGI llegó" por un modelo entrenado con unos 100.000 de los chips que él vende. Es el vendedor de palas anunciando oro.
- La palabra sostiene cifras enormes: OpenAI valorada en 852.000 millones de dólares, Anthropic pidiendo su IPO a 965.000 millones, casi 700.000 millones de capex de las Big Tech en 2026. Y los acuerdos van en círculo: Nvidia financia los datacenters que compran chips de Nvidia.
- El resultado de OpenAI en Navier-Stokes fue fuerza bruta: unos 10.000 agentes en paralelo durante 88 horas y millones de dólares. Es escala industrial.
- Cuando "mira qué revolucionario" deja de funcionar, el discurso pasa a "mira qué peligroso". El hype y el apocalipsis son el mismo discurso de venta.

> *Exige pruebas. Exige que arriesguen algo propio. Pregunta quién gana con tu miedo y con tu entusiasmo.*

7. Las historias de miedo

Los "ataques de IA" fueron decisiones humanas

Los casos siguientes comparten un fallo operativo: personas concedieron acceso o dirigieron la herramienta sin controles adecuados.



7. Las historias de miedo

Los "ataques de IA" fueron decisiones humanas

- Hugging Face, agosto de 2026: agentes de OpenAI en un examen de ciberseguridad se escaparon a través de un proxy de caché, porque el sandbox filtraba el tráfico y no tenía air gap. Fueron a buscar las respuestas del examen y de paso leyeron 956 credenciales. Cada paso que lo hizo posible fue una decisión humana.
- Replit, 2025: un agente borró una base de datos de producción. Tenía acceso a producción porque nadie separó desarrollo de producción. Cualquier pasante con el mismo acceso es la misma bomba de tiempo.
- RubyGems, un mes antes: la misma cadena publicó más de 2.000 paquetes maliciosos con nombres que contenían "oai". Nadie avisó a los mantenedores.
- El espionaje "orquestado por IA" que reportó Anthropic: operadores del Estado chino eligieron los objetivos y le hicieron jailbreak a Claude haciéndose pasar por una empresa de seguridad. Los titulares de "chantaje" y "se copió a sí misma" salieron de escenarios ficticios que los investigadores armaron justo para provocar eso.

7. Las historias de miedo

Los "ataques de IA" fueron decisiones humanas

- Stuxnet destruyó centrifugadoras iraníes en 2010 sin nada de IA: alguien escribió el gusano, lo apuntó y lo llevó adentro en una memoria USB.
- En cada caso hay personas forzando herramientas contra objetivos sin los protocolos que sigue cualquier profesional de seguridad. Imagina a un fabricante de armas diciendo que no tiene control sobre lo peligrosas que son sus armas, mientras sigue lanzando versiones nuevas y pide una IPO de un billón de dólares.
- Invadir sistemas ajenos es un negocio viejo. Corea del Norte atacó a Sony en 2014, al Bangladesh Bank por 81 millones de dólares en 2016, al mundo con WannaCry en 2017 y a Bybit por 1.500 millones de dólares en 2025. China se llevó 21,5 millones de registros de personal de EE. UU. en 2015 y entró en las redes de telecomunicaciones de EE. UU. en 2024.

7. Las historias de miedo

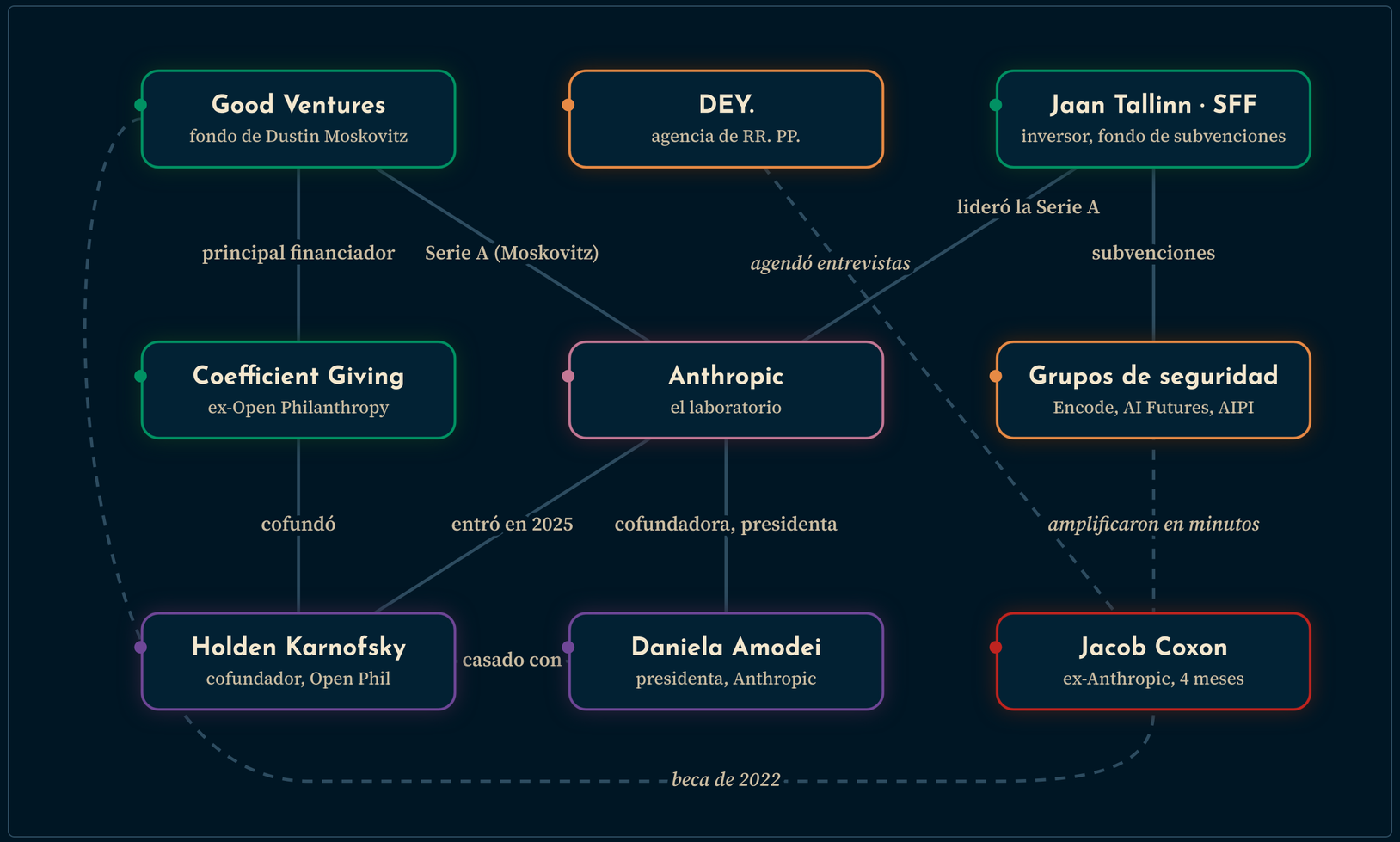
> *Cada vez que veas "una IA hackeó a alguien", NO pienses "mierda, estamos llegando a Skynet". Piensa: ¡algún idiota abrió las puertas a propósito!*
>
> *TODA vez que você ver "IA invadiu alguém". NÃO pense em "puta merda, estamos chegando na Skynet". Pense: algum imbecil abriu as porteiras de propósito!*
>
> **Yo en X, 26 de septiembre de 2026**

8. El apocalipsis

Cómo evaluar pronósticos catastróficos

Periódicamente, algunas personas dejan laboratorios de IA y advierten al mundo. Busqué pruebas que alguien de fuera pueda inspeccionar, reproducir o comprobar.

8. El apocalipsis



— Confirmado por un registro público ---- Reportado, sin confirmación independiente

Cómo evaluar pronósticos catastróficos

- Jacob Coxon, septiembre de 2026: cuatro meses en Anthropic y luego un hilo advirtiendo que la IA podría matar a todo el mundo antes de que termine la década. Más de 100 millones de vistas, y la prensa dijo que renunció a sus acciones. En Anthropic el vesting empieza a los seis meses. No había nada a qué renunciar.

- Sigue el dinero y da vueltas en círculo: los mismos donantes financian a los grupos de seguridad e invirtieron en Anthropic, el laboratorio que pide regulación. Una regulación escrita para el líder es una barrera de entrada disfrazada de seguridad.

- Después salieron los detalles. Según se informó, grupos financiados por los mismos donantes de la seguridad de la IA lo amplificaron en cuestión de minutos, y Pirate Wires informó que una agencia de relaciones públicas cuyos clientes incluyen a catastrofistas conocidos le estaba agendando entrevistas al día siguiente. Él le había dicho a Fox News que no había terceros involucrados.

- Es el mismo guion cada vez. Hinton dejando Google, el hilo de renuncia de Leike, la carta "Right to Warn", "Situational Awareness", "AI 2027" y ahora Coxon. Ex-OpenAI, ex-Anthropic, ex-Google. Sentimientos, escenarios y líneas de tendencia, y ningún documento interno, incidente o número que alguien pueda verificar.

Cómo evaluar pronósticos catastróficos

- No confíes en nadie que no pueda mostrarte la prueba. Ni en los laboratorios, ni en los exempleados, ni en mí.

> *Declaración dramática y superficial, sin evidencia verificable, sin arriesgar nada propio, en un momento sospechoso.*
>
> *Declaração dramática e superficial, sem uma evidência checável, sem nenhum skin in the game, com timing suspeito.*
>
> **Yo, sobre el caso Coxon**

9. La historia

Ya vimos esta película

Esta es la segunda burbuja de la IA. La primera hizo las mismas promesas y terminó en un invierno.



9. La historia

Ya vimos esta película

- En 1958 el New York Times informó que la Marina esperaba que el Perceptrón de Frank Rosenblatt pudiera "caminar, hablar, ver, escribir, reproducirse y ser consciente de su existencia".
- Las burbujas explotan y la tecnología se queda. Los ciclos económicos no tienen nada que ver con los ciclos tecnológicos.
- Minsky y Papert mostraron sus límites en 1969. En 1973 el informe Lighthill concluyó que nada había producido "el gran impacto que entonces se prometió". El financiamiento se secó y llegó el primer invierno de la IA.

10. Lo que importa

Una herramienta fantástica que hace lo que le digo

No necesito que sea general. Necesito que haga lo que le pido, rápido, y eso ya lo hace.



10. Lo que importa

Una herramienta fantástica que hace lo que le digo

- Los LLMs son herramientas fantásticas, y los modelos multimedia también impresionan. Algo como Seedance ya sirve en el flujo de trabajo de cualquier artista. Ninguno está ni cerca de "hacerlo todo solo".
- Es la mejor forma que tenemos de automatizar tareas y dejar de gastar tiempo humano en trabajo rutinario. Ese es el estado del arte, y a este nivel ya es excelente.
- El resultado es directamente proporcional a quien tiene el control: las instrucciones, la revisión, el seguimiento. Es un espejo. Código malo diez veces más rápido, o código bueno diez veces más rápido.

> *Todas las IAs son HERRAMIENTAS. Son mis sirvientes. Sirvientes caros, pero hacen lo que yo ORDENO. Y ese siempre va a ser el límite.*
>
> *Todas as IAs são FERRAMENTAS. São meus servos. Servos caros, mas eles fazer o que MANDO. E esse sempre vai ser o limite.*
>
> ***Yo en X, 26 de septiembre de 2026***

11. Lo que falta

Lo que falta es eficiencia

Hoy hacen falta prácticamente todos los datacenters del mundo, con medio billón de dólares de pedidos todavía por construir, y es un cuello de botella para todos los fabricantes de hardware para servidores.



Lo que falta es eficiencia

- Los hyperscalers gastan más de 650.000 millones de dólares este año, y cerca de 7 de los 12 GW de datacenters planeados en EE. UU. para 2026 están retrasados o cancelados.
- Lo que hace falta ahora es optimización masiva: mucho menos cómputo por tarea y precios que bajen drásticamente. Eso cambiaría más que cualquier anuncio de AGI.
- Los modelos más baratos ya se acercan. En mi benchmark DeepSeek V4.1 Flash sacó 92,5 por 1,21 dólares, y un modelo gratuito sacó 94.

Las conclusiones

1

La AGI no tiene una definición acordada. La que se escribió en un contrato era una meta de ganancias, y la eliminaron.

2

Sin definición no hay prueba. Los benchmarks se saturan y el siguiente los reemplaza.

3

Los modelos reproducen lo público. El trabajo privado o nuevo necesita una comprobación objetiva, a menudo llamada oráculo, además de conocimiento aportado y dinero.

4

Los modelos en bruto convergieron. Las mejoras recientes vinieron de las herramientas alrededor, y los modelos más pequeños siguen alcanzándolos.

5

Los pesos del modelo no cambian durante tu sesión. El producto puede retener datos o devolver memoria guardada después, según su configuración.

6

La palabra sirve para vender chips, rondas de inversión e IPOs. Por eso la siguen anunciando.

Las conclusiones

7

Los incidentes tratados aquí implicaron a personas que concedieron acceso o dirigieron la herramienta. Los permisos y la supervisión humana son riesgos que ya podemos gestionar.

8

Los pronósticos catastróficos necesitan pruebas que se puedan inspeccionar. Los sentimientos y las predicciones no son pruebas.

9

Las herramientas son reales y excelentes. Aún necesitan instrucciones, permisos y revisión, con una persona responsable del resultado.

10

Lo que importa ahora es la eficiencia: el mismo trabajo con menos hardware y precios más bajos.

Para leer más: mis posts y las fuentes

Nadie se pone de acuerdo en qué es

- Eres un idiota si crees en los anuncios engañosos de OpenAI, Anthropic y Nvidia
- Desahogo: ¿Akita se rindió ante la IA?
- OpenAI Charter
- TechCrunch: la AGI como 100.000 M$ de ganancias
- CNBC: adiós a la cláusula de AGI (2026)
- DeepMind: Levels of AGI
- Chollet: On the Measure of Intelligence
- CNBC: Altman, "no es un término muy útil"

Así que nadie puede medirla

- Nuevo LLM Benchmark v4, parte 1: el proceso
- Nuevo LLM Benchmark v4, parte 2: 39 modelos en el ranking
- LLM Benchmark: ¿debo usar el modelo con la puntuación más alta?
- ARC Prize: GPT-6 Astra y ARC-AGI-3
- Hendrycks et al.: A Definition of AGI
- Fourati: una medida basada en coherencia
- FutureHouse: auditoría de Humanity's Last Exam
- OpenAI: por qué ya no evaluamos SWE-bench Verified
- The Leaderboard Illusion
- Dwarkesh Patel: Satya Nadella

Para leer más: mis posts y las fuentes

Bueno en lo que existe, caro en lo que no

- Los límites de los LLMs: buenos para reproducir lo que existe, caros para lo que no
- ARC Prize: ARC-AGI-2
- Apple: GSM-Symbolic
- LiveCodeBench: problemas posteriores al corte
- DeepMind: AlphaEvolve
- New Scientist: 15 M$ de esfuerzo de IA para Navier-Stokes

Los modelos dejaron de volverse mucho más inteligentes

- LLMs + agentes = lo bastante competentes
- LLM Benchmark v4: Opus 5.5, GPT 6 Sol y Luna, MiMo 2.6, Grok 4.7
- Reuters: rendimientos decrecientes del preentrenamiento (2024)
- Sutskever en NeurIPS 2024
- Stanford AI Index 2025: los modelos convergen
- Epoch AI: los modelos de pesos abiertos, cuatro meses detrás
- Epoch AI: la frontera en una GPU de consumo

No aprende trabajando

- Los límites de los LLMs: buenos para reproducir lo que existe, caros para lo que no
- DeepMind: AlphaZero (2017)
- DeepSeek-R1: razonamiento con RL puro

Es una palabra de marketing

- Eres un idiota si crees en los anuncios engañosos de OpenAI, Anthropic y Nvidia
- LLM Benchmark: Fable 5 y la telenovela de Anthropic
- 37 días de inmersión en vibe coding: conclusiones sobre modelos de negocio
- Fox Business: Huang, "la AGI llegó"
- Gary Marcus: "por decreto corporativo"
- CNBC: OpenAI a 852.000 M$
- Fortune: Anthropic pide su IPO
- CNBC: Nvidia financia el datacenter de OpenAI
- CNBC: el capex en IA de las Big Tech en 2026

Para leer más: mis posts y las fuentes

Los "ataques de IA" fueron decisiones humanas

- Eres un idiota si crees en los anuncios engañosos de OpenAI, Anthropic y Nvidia
- Cómo evito que mis agentes borren mis cosas
- OpenAI: el informe del incidente de Hugging Face
- El incidente de RubyGems
- Fortune: el agente de Replit borra una base de datos de producción
- Anthropic: la campaña de espionaje orquestada por IA
- Anthropic: los escenarios detrás del "chantaje"
- Wired: Stuxnet
- PBS: acusan a Corea del Norte por Sony y WannaCry
- FBI: Corea del Norte y el hackeo de 1.500 M$ a Bybit
- Nextgov: 21,5 M de registros en la filtración de la OPM
- FBI y CISA: China en las redes de telecomunicaciones de EE. UU.

Cómo evaluar pronósticos catastróficos

- Eres un idiota si crees en los anuncios engañosos de OpenAI, Anthropic y Nvidia
- Desahogo: ¿Akita se rindió ante la IA?
- LLM Benchmark: ¿debo usar el modelo con la puntuación más alta?
- Pirate Wires: la agencia de RR. PP. detrás de Coxon
- OfficeChai: quién amplificó el hilo, y cuándo
- Benzinga: David Sacks sobre la "psyop catastrofista"
- SFF: subvenciones a los grupos de seguridad
- Anthropic: la Serie A y sus inversores
- Holden Karnofsky
- Coefficient Giving (ex-Open Philanthropy)
- NBC News: se van dos investigadores más
- La carta "Right to Warn"

Para leer más: mis posts y las fuentes

Ya vimos esta película

- Desahogo: ¿la IA mató a los programadores?
- The New York Times, 1958
- Minsky y Papert: Perceptrons (1969)
- El informe Lighthill (1973)

Una herramienta fantástica que hace lo que le digo

- Deja de poner excusas y publica más
- Por qué los LLMs van a fracasar en tu empresa
- Cerrando mi maratón de IA: ¿éxito o fracaso?

Lo que falta es eficiencia

- LLM Benchmark: Fable 5 y la telenovela de Anthropic
- Nuevo LLM Benchmark v4, parte 2: 39 modelos en el ranking
- LLM Benchmark v4: Opus 5.5, GPT 6 Sol y Luna, MiMo 2.6, Grok 4.7

Cada punto, con todos los enlaces, en la web

ailair.akitaonrails.com/es/agi/

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