

El mapa del territorio

Toda la inteligencia artificial del mundo cabe en una imagen: una pila de seis capas. La recorreremos de abajo hacia arriba, porque el poder se acumula donde nadie mira.

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

INTRODUCCIÓN

Casi todos quieren ir por delante de la inteligencia artificial. Muy pocos quieren entenderla, que es la única forma de conseguirlo.

Entender la IA ocurre en tres frentes a la vez. La economía: quién pone el dinero y quién lo recoge. La geopolítica: quién controla las piezas que casi nadie sabe fabricar. Y lo técnico: qué hace la máquina, de verdad, cuando uno le habla.

Esta serie se propone dejar los tres claros para un lector no especialista. Todo empieza por el mapa: dónde vive cada cosa, y quién manda en cada parte.

La intuición es cómoda: la IA sería el asistente con el que uno conversa. Es apenas la capa de arriba.

La pila de seis capas

6 Aplicaciones

← *lo que ves*

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

Debajo hay cinco capas más, y toda respuesta las atraviesa hacia abajo, hasta la corriente que la enciende. Conviene leerlas al revés de como se ven, de abajo hacia arriba.

En cada capa mandan tres o cuatro nombres. El resto del mundo les alquila.

Las seis capas no pesan igual. Cuanto más abajo, menos jugadores, y más difíciles de reemplazar.

Mucho arriba, poquísimo abajo



Miles de aplicaciones se apoyan sobre unos pocos modelos; esos modelos se entrenan casi todos sobre los chips de una empresa; esos chips salen, en su mayoría, de una sola fábrica.

La regla del sótano: cuanto más raro es fabricar algo, más manda quien lo fabrica.

1

lo físico

El sótano

Antes de ser lógica, la IA es materia: hace falta con qué alimentarla y con qué construirla.

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

Capa 1 • Energía. Uno diría que un modelo vive en internet, en ninguna parte. Vive en galpones de máquinas que consumen como una ciudad.

La corriente

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 **Energía**

1,5%

de la electricidad del mundo iría ya a los centros de cómputo, y podría duplicarse hacia 2030.

Entrenar un modelo grande es, antes que nada, un problema de potencia: los datacentros nuevos se miden en gigavatios (GW), la escala de una ciudad. El primer límite de la IA es hoy la red eléctrica.

Capa 2 • Silicio. Esa potencia mueve un chip, y el chip nace de la arena: silicio purificado casi hasta el átomo.

La fábrica

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

~90%

de los chips de vanguardia saldría de una sola empresa, la taiwanesa TSMC, en buena parte de una sola isla.

Grabar circuitos de pocos nanómetros exige máquinas de litografía (luz ultravioleta extrema, EUV) que fabrica una única compañía en el planeta: ASML. Dos empresas sostienen, en silencio, todo lo que viene arriba.

2

el cómputo

El motor

Con corriente y silicio ya se puede construir la máquina que calcula, y alquilarla. Dos capas: el chip y la nube.

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

Capa 3 • Aceleradores. Un modelo hace una sola cosa, billones de veces: multiplicar números en paralelo. El chip hecho para eso se llama acelerador.

El peaje

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

>80%

del entrenamiento de IA correría sobre aceleradores de una sola empresa: NVIDIA.

Un procesador común (CPU) resuelve tareas variadas de a una; el acelerador repite la misma cuenta a gran escala, que es lo que un modelo pide. Los que escapan fabrican el suyo (Google, TPU; Amazon, Trainium). El resto paga el peaje.

Capa 4 • Cómputo. Un acelerador de punta cuesta como un carro; entrenar un modelo pide decenas de miles. Casi nadie los compra: los alquila por horas.

El arriendo

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

US\$350

mil millones habrían gastado en infraestructura, solo los cuatro grandes de la nube, en 2025; 2026 apunta más alto.

Eso es una nube: chips de otro, en el galpón de otro, cobrados por hora. Arriba, los gigantes (Amazon, Microsoft, Google); debajo, los neoclouds (CoreWeave, Nebius), que solo alquilan GPUs. Alquilar cómputo es ya un negocio del tamaño de una industria pesada.

3

la inteligencia

Lo visible

Recién en la cima aparece lo que llamábamos IA. Y ahí arriba solo hay dos formas de ganar dinero.

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

Con la máquina construida y encendida, queda venderla. Casi todo el sector está en una de dos jugadas.

Los dos juegos

1

Vender la inteligencia

Se cobra por uso, en tokens.

los laboratorios de modelos

2

Vender productos

Se cobra por el producto que la usa.

las aplicaciones

Uno vende el motor; el otro, el vehículo que lo lleva puesto.

Juego 1 • Los modelos. Los laboratorios de frontera venden inteligencia cruda, cobrada por token: la unidad mínima de texto, más o menos una sílaba.

Vender inteligencia

6 Aplicaciones	ESTADOS UNIDOS	CHINA
5 Modelos	OpenAI	DeepSeek
4 Nubes	Anthropic	Alibaba · Qwen
3 Chips	Google DeepMind	Moonshot · Kimi
2 Silicio	xAI	MiniMax
1 Energía	Meta	

Entrenar cada modelo cuesta cientos de millones de dólares. Capital enorme y margen frágil, sobre una apuesta: que la inteligencia siga siendo escasa.

Fuente: Epoch AI (estimaciones de cómputo)

Juego 2 • Las aplicaciones. El resto del sector no entrena nada: toma un modelo ajeno y construye encima un producto que alguien pague.

Vender productos

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

miles

de productos, contruidos sobre un puñado de modelos ajenos.

Es barato empezar y rápido crecer, pero el motor es alquilado: el margen lo fija la capa de abajo, y la competencia copia el producto en meses. Su apuesta es la contraria: que la inteligencia se abarate.

Una de las dos apuestas ya está perdiendo, y el precio lo dice.

El precio de la inteligencia

~10×

más barata cada año: el costo de un token, para una misma capacidad, habría caído cerca de un orden de magnitud al año.

Si la inteligencia se abarata así de rápido, vender el modelo empieza a parecerse a vender electricidad: un servicio básico, difícil de diferenciar. La escasez, de momento, no está en el modelo.

Las dos jugadas de arriba descansan sobre apuestas opuestas, y ninguna es segura todavía.

¿Quién se queda con el valor?

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía

El laboratorio gana si la inteligencia sigue cara; la aplicación, si se abarata. Mientras esa disputa se resuelve arriba, abajo el sótano cobra en firme: la energía y los chips se pagan pase lo que pase.

El valor visible cambia de dueño cada temporada. El valor de abajo, casi nunca.

Sobre el mapa cae, además, una frontera política que lo parte de lado a lado.

La grieta

EE. UU.	China
6	6
5	5
4	4
3	3
2	2
1	1

capa 3: el corte de los controles

Desde 2022 Estados Unidos prohíbe venderle a China los aceleradores más avanzados, y apunta justo a la capa 3, donde su ventaja es mayor. China responde con modelos que rinden con menos silicio de punta (DeepSeek, Qwen).

Presuntamente, la carrera se juega tanto en los laboratorios como en las aduanas.

El sótano manda

Este mapa es el índice de la serie: cada pieza que venga se asomará a una celda, un modelo, un chip, una nube, una frontera. Vuelva a él cuando se pierda.

En toda fiebre del oro, el oro cambia de manos cada temporada; las palas, no.

6 Aplicaciones

5 Modelos

4 Nubes

3 Chips

2 Silicio

1 Energía