Nota de investigación · Noviembre de 2026

El cómputo como consenso

IA × cripto × chips en 2026, a partir de Pearl

Pearl · Trabajo útil · DePIN y agentes · Economía de los ASIC · Nodos de proceso · Presupuestos por etapas

Tan Shuai
tanshuai.com · i@tanshuai.com



El teorema se sostiene; el incentivo, no

01 Pearl funciona como ingeniería, aún no como trabajo útil

Unos 90 EH/s de multiplicación de matrices enteras protegen la cadena, pero solo cerca del 1 % de los bloques lleva trabajo real de IA.

02 La demanda de pago está en la inferencia, los datos y la privacidad

Bittensor, unos 70 millones de dólares al año de ingresos verificables; Grass, 17 millones en el S1 de 2026; TEE en producción; pagos de agentes, unos 5000–11.000 dólares al mes.

03 Un ASIC solo compensa con un bote grande, reglas estables y un segundo mercado

Su techo de beneficio es R × (1 − 1/k). Hoy ninguna red ofrece las tres cosas análisis .

≈90 EH/s

Pearl el 13 de noviembre de 2026: el rendimiento entero de unas 122.000 H100

≈1 %

de los bloques de Pearl llevan trabajo real de IA (Blockworks)

4–11×

las emisiones de Bittensor frente a sus ingresos verificables

EL PANORAMA

IA × cripto en seis sectores: dónde hay demanda de pago real

Cadenas de trabajo útil — 01

REAL PERO PEQUEÑO

≈70 M$ al año

ingresos externos verificables de Bittensor

Señal: emisiones 4–11× los ingresos; Pearl ≈1 % de bloques con IA

Los mecanismos funcionan; la utilidad necesita pagadores

Cómputo descentralizado — 02

REAL PERO PEQUEÑO

2,35 M$ al año

comisiones on-chain de Render, las mayores verificables

Señal: valor de mercado ≈360× las comisiones

Sobre todo reventa de inferencia genérica

Entrenamiento e inferencia — 03

BENEFICIO EN EL CAPITAL

100+ M$ al año

Prime Intellect, una empresa de capital en GPU centralizadas

Señal: la mainnet de entrenamiento de Gensyn no está activa

El avance es real; el beneficio va al capital

IA verificable — 04

INCIPIENTE

398

usuarios de pago de la nube TEE de Phala, finales de 2025

Señal: sin zkML de pago declarado

TEE en producción; zkML incipiente

Pagos de agentes — 05

ANTES DE LA DEMANDA

5–11 mil $ al mes

pagos x402 iniciados de verdad por agentes

Señal: ≈200 M de transacciones, casi todo señalización

Primero llegaron los raíles; la demanda, no

Redes de datos — 06

COMPRADORES REALES

17 M$

ingresos por datos de Grass, primer semestre de 2026 (autodeclarados)

Señal: valor de mercado ≈10× los ingresos anualizados

Uno de los pocos sectores con compradores reales

> El subsidio es combustible para el arranque en frío, no un modelo de negocio. Lo cripto aporta distribución y liquidación; el capital se lleva los beneficios, al menos en 2026.

Una multiplicación-acumulación es un hash

Del teorema a la mainnet

Komargodski y Weinstein (abril de 2025): una prueba de trabajo útil para cualquier producto de matrices con un sobrecoste de 1 + o(1). La mainnet de Pearl arrancó el 27 de abril de 2026.

Cómo funciona

Los mineros multiplican matrices con ruido, convierten el acumulador de cada tesela en un boleto de lotería mediante un hash y prueban la tesela ganadora con una prueba Plonky2 de menos de 60 KB.

La unidad

1 H es una multiplicación-acumulación INT7 × INT7 → INT32 acreditada: el presupuesto de seguridad de una cadena, valorado en rendimiento tensorial.

PEARL EL 13 DE NOVIEMBRE DE 2026

89,77 EH/s

hashrate de la red (nodo completo independiente)

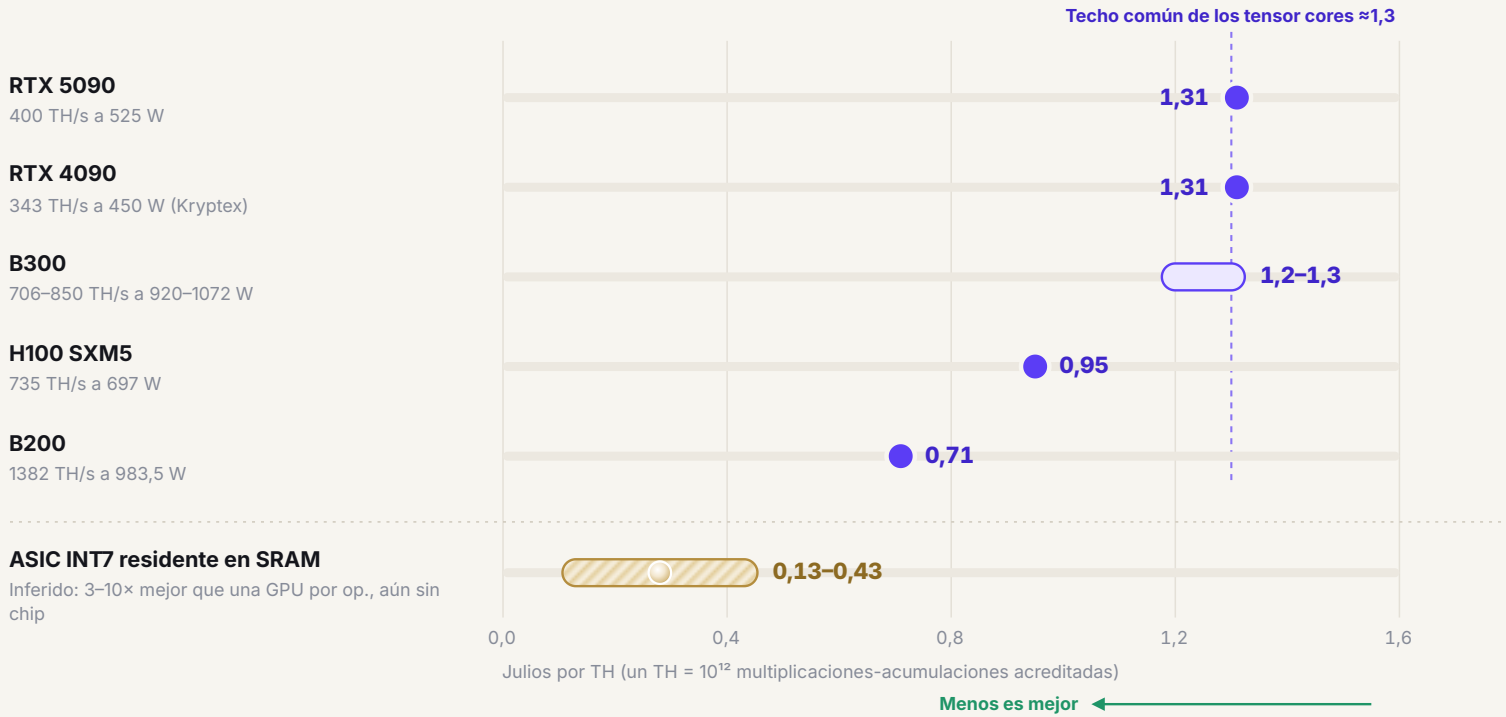
≈500 M$

valor de mercado; totalmente diluido, unos 2,8 mil millones de dólares

+78 %

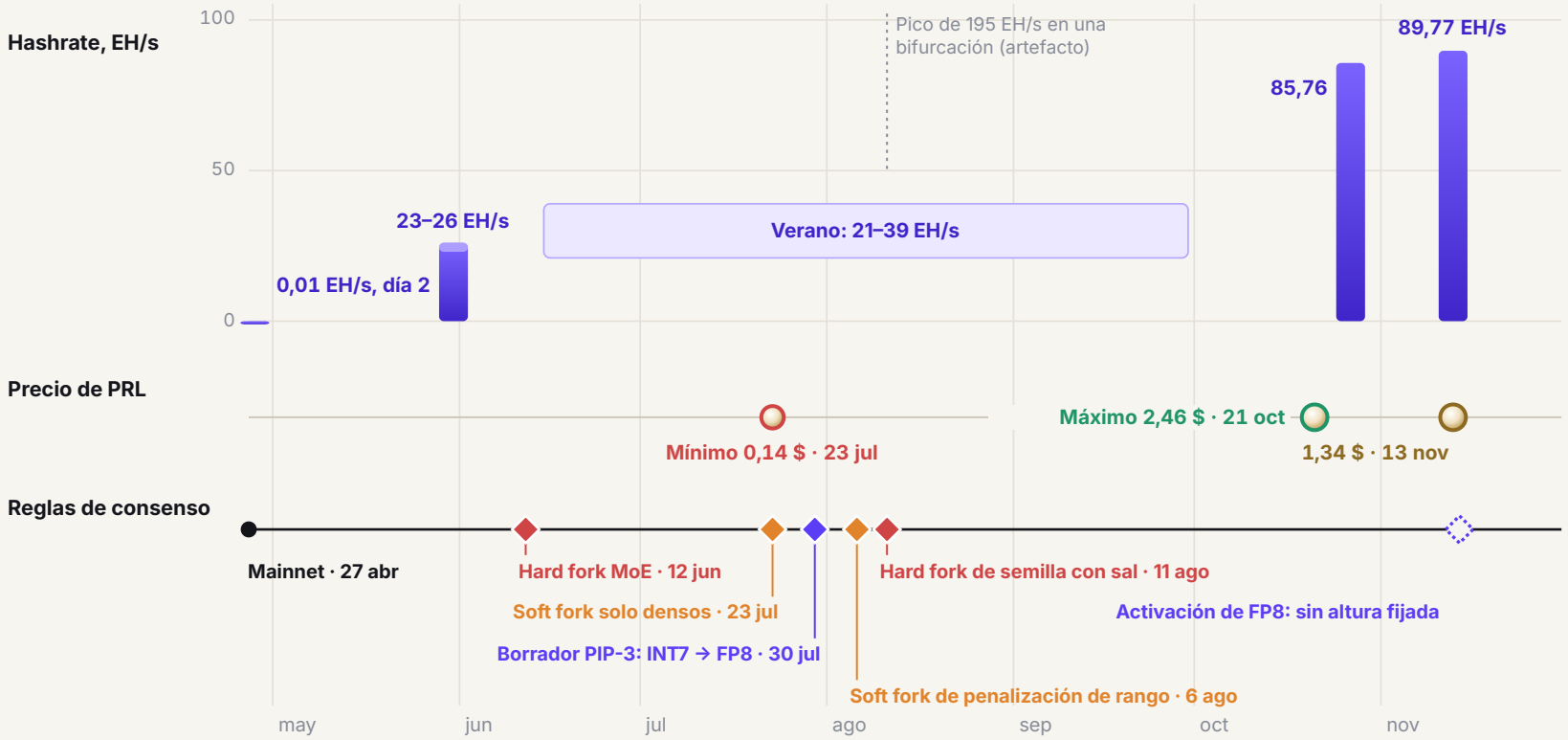
de nueva oferta en los próximos 12 meses

En el trabajo entero de Pearl, las GPU comparten un techo energético



> Las tarjetas gaming y la B300 coinciden en torno a 1,3 J/TH; incluso la B200 solo llega a 0,71. Un chip que solo multiplique enteros de 7 bits desde SRAM podría plausiblemente ser 3–10 veces mejor **análisis**.

El hashrate sigue al precio; las reglas no dejan de cambiar



Unos 120 millones de monedas, el 32 % de la oferta actual, se minaron en los seis primeros días. Una RTX 5090 gana ahora unos 5–6 dólares al día; alquilar tarjetas solo para minar pierde dinero.

El teorema se sostiene; el incentivo, no

LO QUE CUESTA EL TRABAJO ÚTIL

Lado de la inferencia: poco sobrecoste

Pearl declara un 5,08 % más de tiempo para Llama 70B en cuatro H200; una prueba independiente midió un 9,2 % menos de rendimiento en una GPU Hopper.

LO QUE GANA EL TRABAJO ÚTIL

Lado de la minería: muy poco hashrate

Una tarjeta que sirve inferencia mina a cerca del 27 % de un minero dedicado en el prefill, al 5,4 % en chat y a cerca del 1 % cuando domina la decodificación.

Así que los mineros racionales multiplican matrices aleatorias. La utilidad es un problema comercial fuera de la cadena: alguien tiene que pagar la inferencia.

Cuatro bifurcaciones en dos meses, y después FP8

12 de junio
Hard fork de mezcla de expertos (bloque 71.935)

23 de julio
Soft fork solo para modelos densos (bloque 91.630)

6 de agosto
Soft fork de penalización de rango: algunas configuraciones pierden el 50–87,5 % del hashrate efectivo

11 de agosto
Hard fork de semilla con sal: los mineros antiguos solo producen shares inválidas

Sin fijar
PIP-3: de INT7 a FP8, solo se admiten Hopper y Blackwell

2 de 3
útil para la IA, verificable bit a bit, neutral respecto al hardware: un formato numérico logra como mucho dos

13 bits
conserva la acumulación FP8 de Hopper, con truncamiento: no reproducible entre generaciones

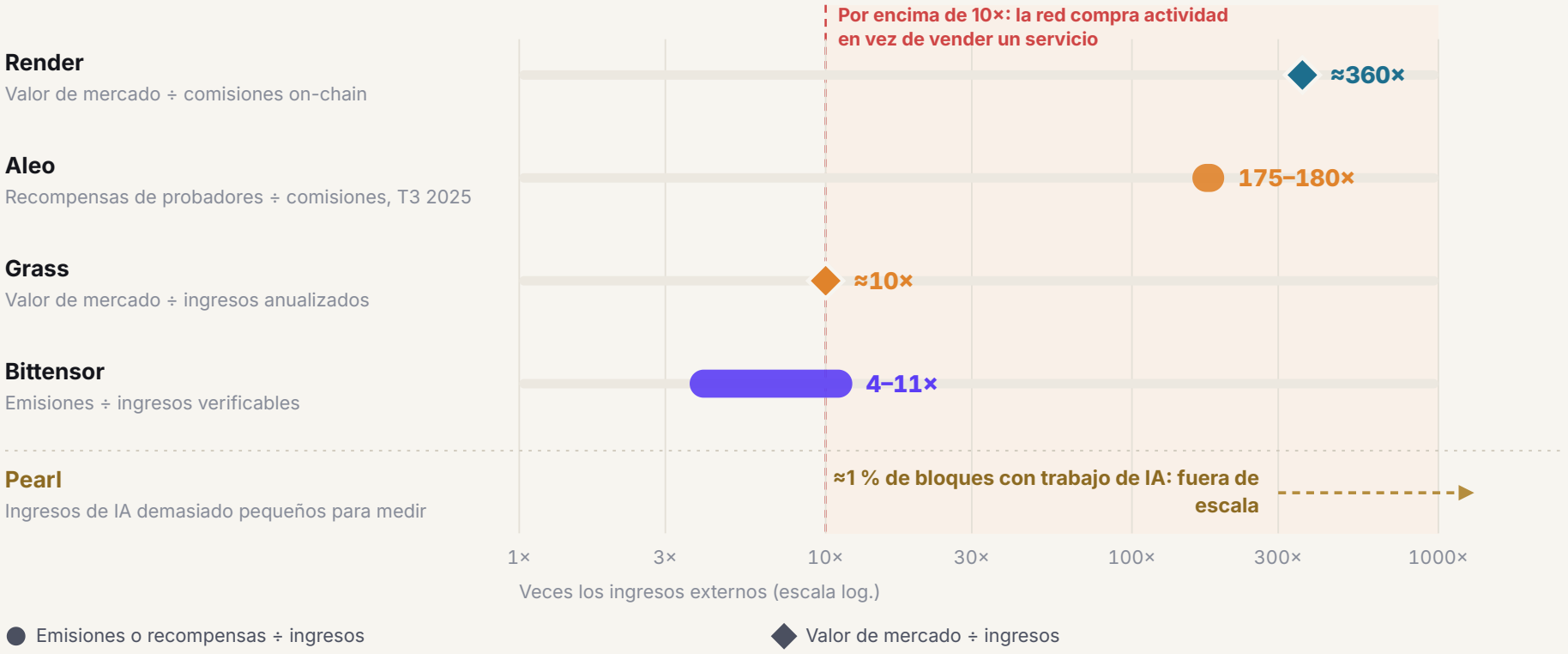
≈82 %
del hashrate en tarjetas de consumo que PIP-3 no admitiría (modelo de baja confianza)

Quien controla la función de trabajo controla el destino de quienes invierten en hardware análisis **.**

Ni las matrices ni el conocimiento cero frenan a los ASIC

Red	Qué pasó	Qué enseña
Kaspa	Un producto de matrices 64 × 64 entre hashes; ASIC a los 16 meses, unas 1000× la eficiencia de una GPU; el mejor gana hoy 0,86 dólares al día	Una matriz no es ni útil ni resistente a los ASIC
Aleo	Un rompecabezas SNARK sintético; máquinas dedicadas a los cinco meses; recompensas de probadores 175–180× las comisiones	El trabajo sintético vive del subsidio
Qubic	Dice entrenar IA; su cómputo minó Monero (51 % declarado) y ahora Dogecoin	Una PoW de propósito general puede ser capturada desde fuera
Bittensor	Puntuación ponderada por stake; unos 70 millones de dólares al año de ingresos verificables frente a más de 300 millones de emisiones	La única con ingresos externos a escala
Monero, Sia, Ethereum	Las bifurcaciones inutilizaron ASIC; tras el Merge, menos del 20 % de los mineros encontró otro trabajo	Sobrevive el hardware con un segundo mercado

El múltiplo de subsidio es el chequeo de salud más honesto



> Por encima de unas 10 veces los ingresos externos, una red compra actividad en vez de vender un servicio. Los ingresos de IA de Pearl son demasiado pequeños para situarlos en la escala.

Demanda real pero pequeña, y el valor se va al capital

- **Mercados de cómputo**

  Comisiones on-chain de Render, unos 2,35 millones de dólares al año; la H100 de Akash a 2,73 dólares la hora supera a AWS y Google, pero no a Vast.ai ni a RunPod.

- **Entrenamiento**

  Modelos de 8000–40.000 millones de parámetros entrenados por internet; INTELLECT-3 volvió a 512 H200 con InfiniBand.

- **IA verificable**

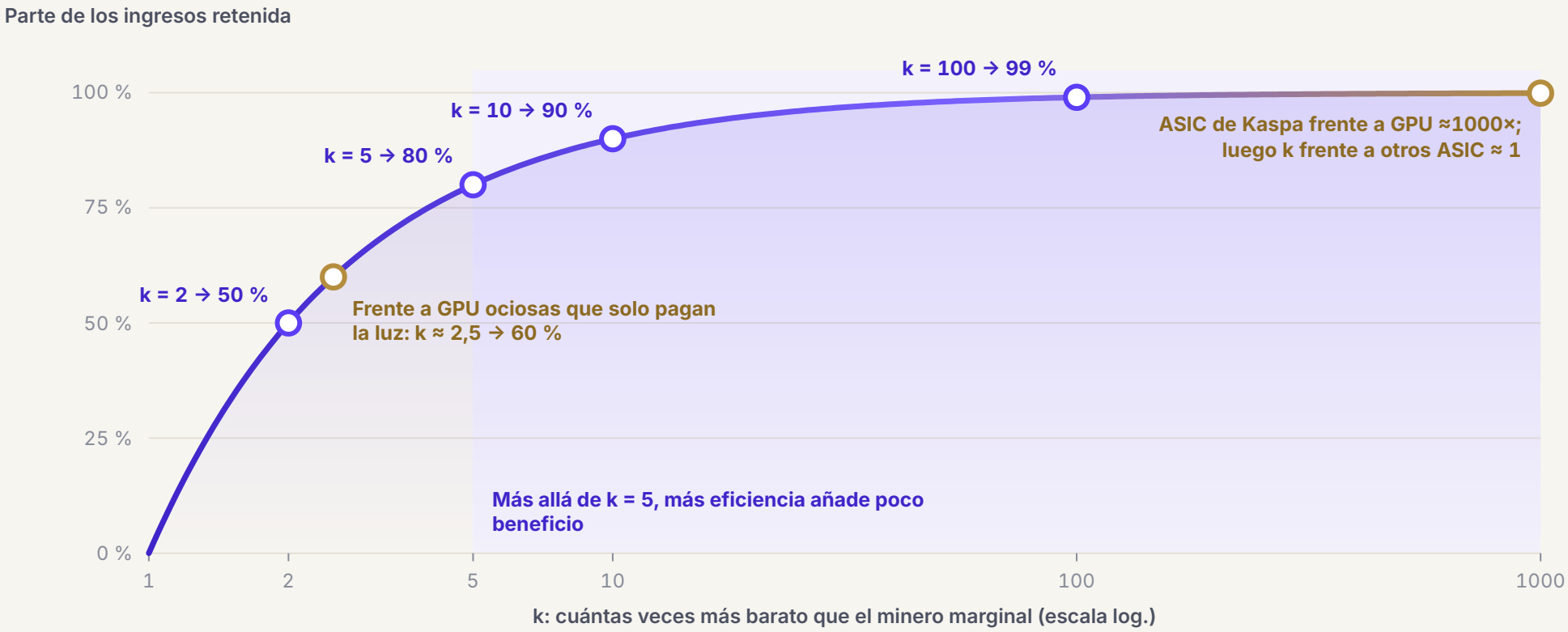
  Los TEE añaden menos del 7 % y funcionan en producción; TEE.fail rompió la atestación con un interposer de 1000 dólares; zkML prueba GPT-2 a 174 tokens por minuto.

- **Agentes y datos**

  Flujos x402 iniciados por agentes, unos 5000–11.000 dólares al mes; Grass, 17 millones de dólares de ingresos por datos en el S1 de 2026.

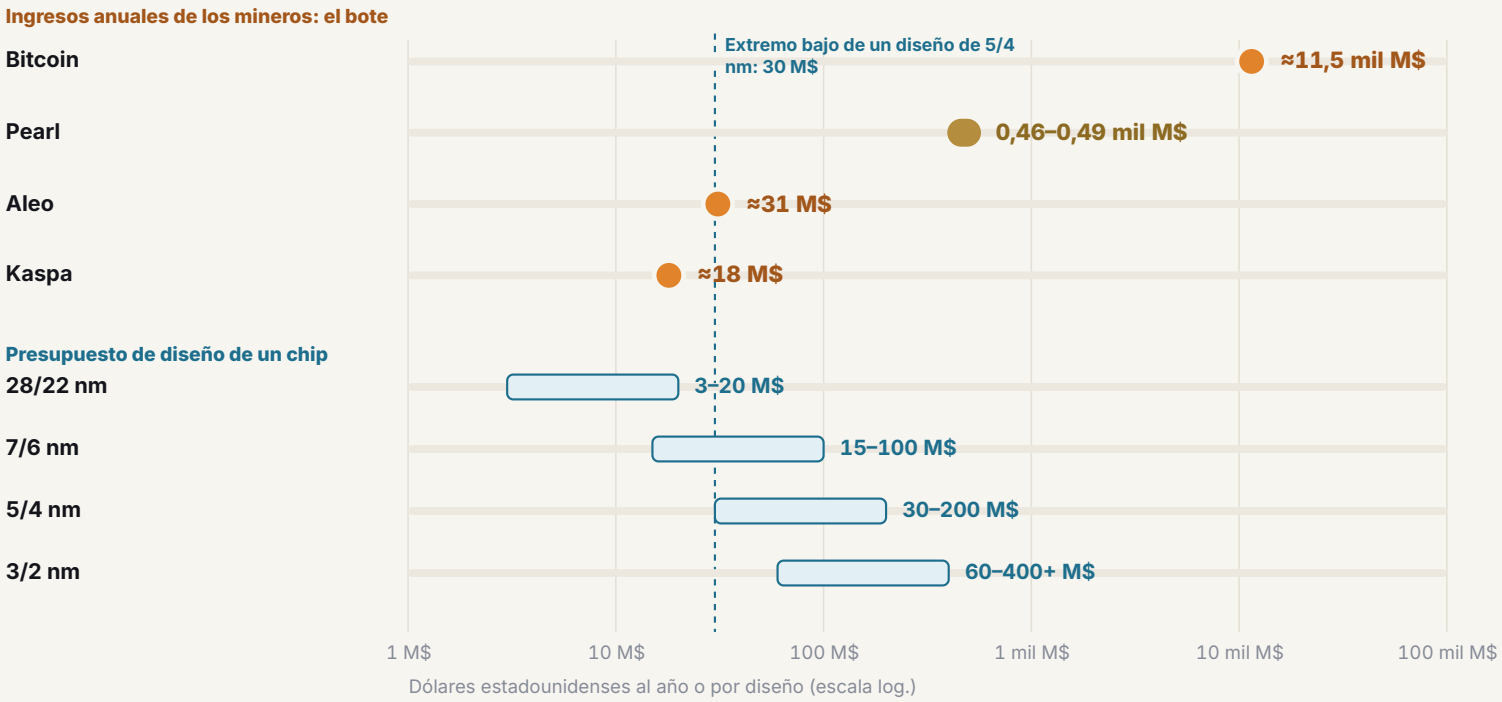
> **Cuatro empresas de capital sobre infraestructura centralizada valen unos 8,5 mil millones de dólares, más o menos lo mismo que todos los tokens de IA de capa de protocolo y de aplicación.**

Por qué la eficiencia deja de pagar a partir de k = 5



Techo de beneficio ≈ R × (1 − 1/k). Cuando los ASIC pasan a ser los mineros marginales, k vuelve a medirse frente a otros ASIC; frente a GPU ociosas que solo pagan la luz, k ronda 2,5 estimación.

La mayoría de los botes son menores que el presupuesto de diseño de un chip



> Hashrate de amortización h* = NRE ÷ [R × (1 − 1/k) × T × p]. Para Pearl con 20 millones de dólares de NRE y k = 5: unos 4,9 EH/s durante un año si las reglas se mantienen, 9,8 EH/s si se mantienen con probabilidad del cincuenta por ciento **estimación**.

Cinco tipos de chip, cinco respuestas



Chip de minería PoUW al estilo de Pearl

AHORA NO

k: sin HBM ni unidades generales: 3–10× por op. (inferido)

Segundo mercado: inferencia de IA, si es GEMM INT8/FP8 general

Apostar por un chip de inferencia que pueda minar, no por una pieza INT7 dedicada



Chip de minería SHA-256 de bitcoin

MANTENERSE FUERA

k: ≤1 frente a los líderes

Segundo mercado: casi ninguno

Solo con energía propia, un formato de nicho o como IP



Chip de entrenamiento de IA

IP, NO CHIPS

k: 1,1–1,8× a igual precisión

Segundo mercado: —

Deciden el encapsulado y la HBM; hacer IP, chiplets, servicios



Chip de inferencia de IA

CONSTRUIR

k: renunciar a la generalidad: 10–20× declarado

Segundo mercado: es el mercado principal

Firmar al cliente y luego elegir el grado de atadura



Acelerador de pruebas ZK

ESPERAR Y CUBRIRSE

k: 10–100× las GPU, declarado

Segundo mercado: limitado fuera de ZK

Motores MSM y NTT programables; primero FPGA



Acelerador FHE

POSICIONARSE PRONTO

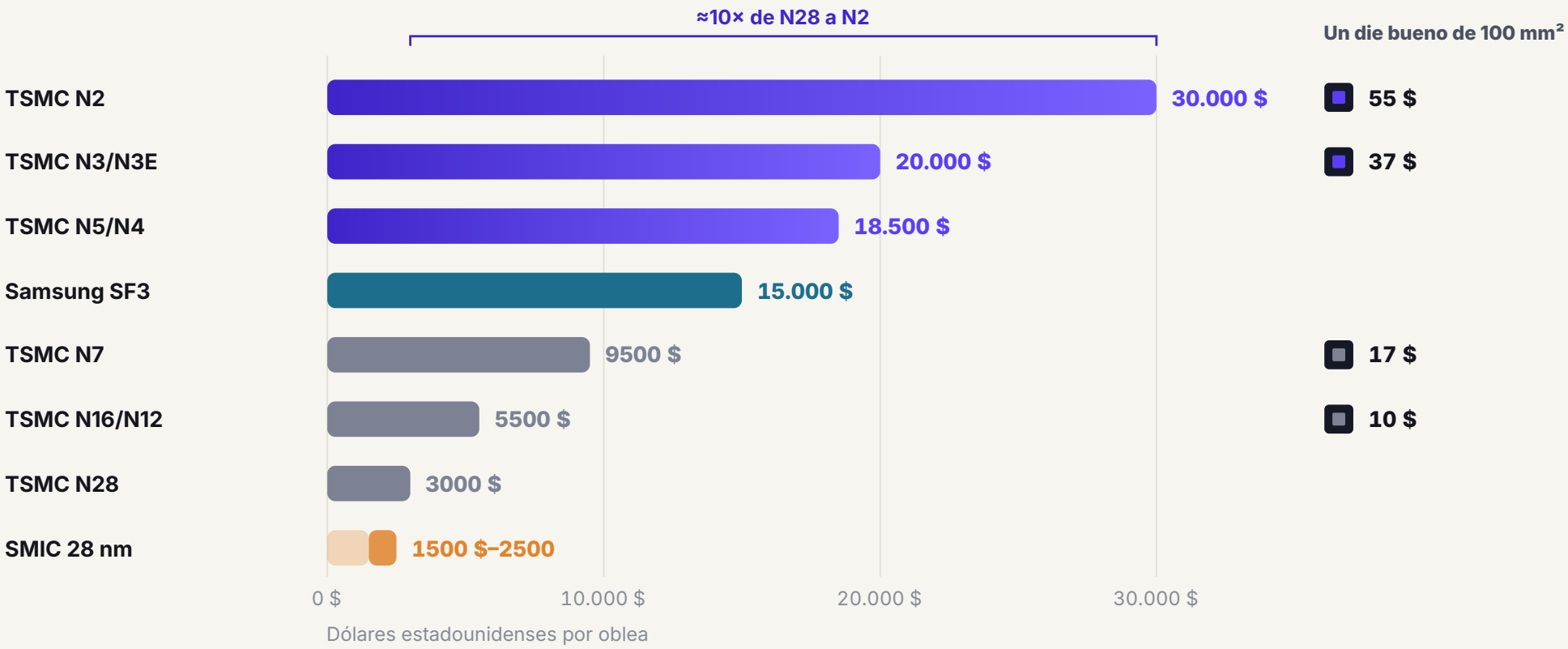
k: 1074–5547× un Xeon (Heracles)

Segundo mercado: computación cifrada, IA privada

Una ventana 2027–2029; validar con activos ligeros

> Hoy ninguna red reúne un bote grande, reglas estables y k por encima de 1: la k de bitcoin es de los líderes, las reglas de Pearl son las menos estables y los botes de Kaspa y Aleo son demasiado pequeños.

Una oblea de vanguardia cuesta diez veces una madura



> Que decida el mayor coste a lo largo de la vida útil: la energía en los chips de cómputo continuo, el encapsulado y la HBM en los de entrenamiento (63 % del coste de una H100), la elección entre SRAM y HBM en los de inferencia.

Un nodo para cada chip y cada equipo

Chip	Equipo pequeño	Empresa a escala	Vía de China continental
Chip de minería PoUW	MPW en 28 o 16/12 nm; primer volumen en N7/N6	N3/N3E o Samsung SF3	SMIC N+1/N+2 sirve, con un 40–50 % más por die
Chip de minería SHA-256	No perseguir la frontera	Clase 3 nm	Difícil igualar la eficiencia del nodo puntero
Chip de entrenamiento de IA	IP, chiplets, servicios de diseño	N3P y luego N2/A16 con CoWoS y HBM	SMIC N+2/N+3 con HBM nacional
Chip de inferencia de IA	Centrado en SRAM en N6/N7 o N4/N5/SF4	N3/N2 con HBM, o chiplets de SRAM	SMIC N+2 con LPDDR o GDDR
Aceleradores ZK y FHE	Primero FPGA o MPW en 16/12 nm	N5/N3; FHE con HBM	ZK viable; FHE limitado por la HBM

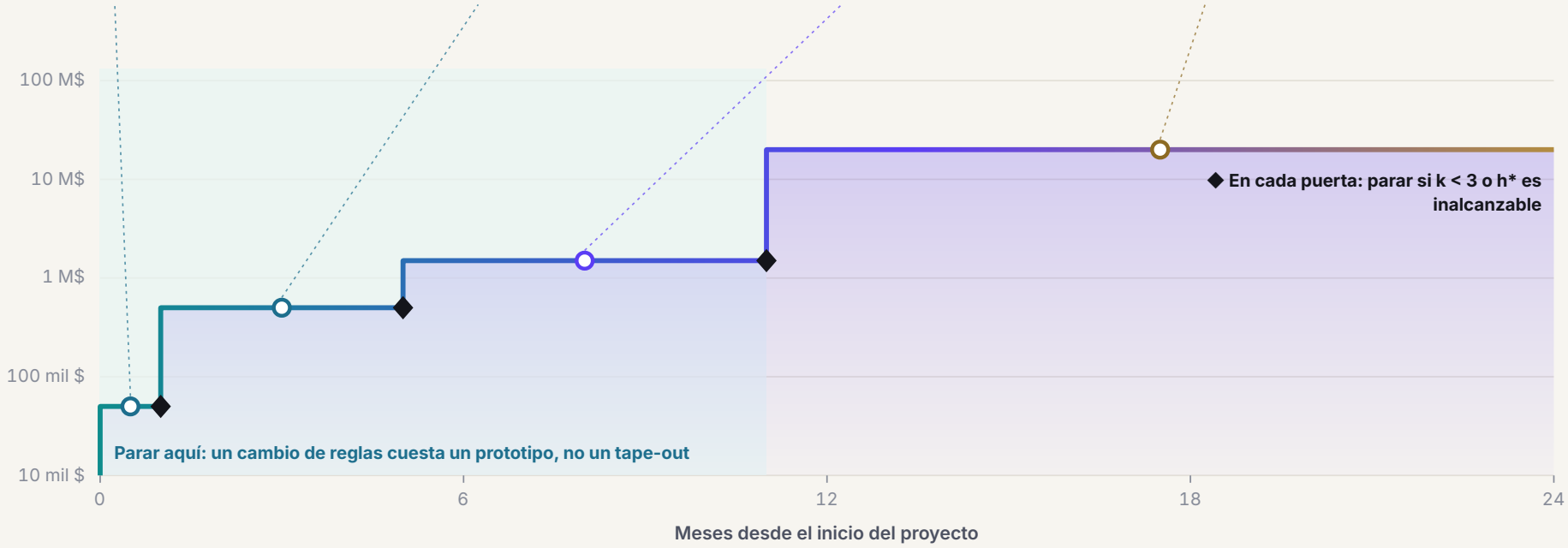
Pruebas antes que máscaras

Puerta 0 · Modelo
≈50 mil $ · semanas
J/op y $/op a partir de la aritmética, la SRAM y la E/S

Puerta 1 · FPGA
≈0,5 M$ · meses
La carga real de extremo a extremo: shares o tokens

Puerta 2 · MPW
≈1,5 M$ · mes 11
El silicio medido sustituye al modelo

Puerta 3 · Máscaras y primer lote
≈20 M$ · mes 24
Máscaras N7 ≈3,5 M$ dentro de un diseño de 15–100 M$



Un modelo, una FPGA y una MPW llegan a silicio medido por unos 1,5 millones de dólares en once meses; el juego de máscaras y el primer lote cuestan más de diez veces más. Hay que parar antes de las máscaras hasta que estén fijados el comprador o la versión del protocolo.

Lo cripto como primer cliente, la IA como segundo mercado

Chips a medida

- La demanda es abrumadoramente de IA: Broadcom, unos 58 mil millones de dólares en semiconductores de IA en el ejercicio 2026
- Lo aprendido en chips de minería se transfiere: matrices de baja tensión, potencia de kilovatios, tape-outs rápidos
- Ganar con una vía propia al mercado, no con los canales de la minería

Desarrollo de IA

- Las horas son un commodity: 25–49 dólares la hora en Clutch
- El valor está en las juntas: kernels para chips a medida, minería dentro del servicio de inferencia
- Inferencia verificable: TEE más comprobaciones puntuales con ZK

Sistemas IA × cripto

- Verificación por capas: TEE o muestreo, luego disputas o comprobaciones puntuales con ZK
- Las barreras son la distribución y la liquidez, no el coste
- Firmar al comprador de cómputo antes que el token

Tres caminos, cada uno con un primer paso pequeño

Camino	Primer paso
A. Un acelerador pensado para inferencia que también pueda minar	Un estudio de cuatro semanas de J/op y $/op de una matriz INT8/FP8 residente en SRAM frente a la RTX 5090 y la H100; después, FPGA o MPW en nodo maduro
B. Servicios de silicio a medida: primero cripto, después IA	Convertir el modelo de NRE y de punto de equilibrio de este informe en estudios de viabilidad de pago, con hitos y una versión de protocolo fija
C. Una red nativa de IA con un comprador antes que un token	Una nota de diseño sobre verificación por capas y una aritmética canónica neutral respecto al fabricante, redactada con un comprador de inferencia concreto

Como mucho tres caminos, cada uno probado antes de comprometer dinero análisis.

Qué vigilar y qué preguntar

INDICADORES

En cualquier momento	La altura de activación de FP8 en Pearl, y si la cadena se divide
Mensual	La proporción de bloques de Pearl con trabajo de IA
Trimestral	Ingresos verificables de Bittensor frente a emisiones; flujos x402 iniciados por agentes
Enero de 2027	La subida de precios del 3–6 % de TSMC en 2 nm y 3 nm, según la prensa
T1 2027	FP4 en Pearl, según lo contado en el AMA de los fundadores
Abril de 2028	El próximo halving de bitcoin

PREGUNTAS QUE HACER

01 ¿Quién controla la función de trabajo?

02 ¿Cuál es el múltiplo de subsidio?

03 ¿Cuánto cuesta la siguiente puerta, y qué le hace parar?

04 ¿Cuál es el hashrate de amortización, con p?

05 ¿Cuál es el segundo mercado del chip?

06 ¿Quién es el primer cliente de pago, con nombre?

Son áreas que podemos explorar juntos, mediante intercambio y colaboración técnicos: chips a medida, desarrollo de IA y sistemas IA × cripto.

i@tanshuai.com · tanshuai.com