Ghi chú nghiên cứu · Tháng 11 năm 2026

Tính toán là đồng thuận

AI × Crypto × Chip năm 2026, khởi đầu từ Pearl

Pearl · Công việc hữu ích · DePIN và tác tử · Kinh tế học ASIC · Tiến trình · Ngân sách theo giai đoạn

Tan Shuai
tanshuai.com · i@tanshuai.com



KẾT LUẬN

Định lý đứng vững; động lực thì không

01 Pearl thành công về kỹ thuật, nhưng chưa phải công việc hữu ích

Khoảng 90 EH/s phép nhân ma trận số nguyên bảo vệ chuỗi, nhưng chỉ khoảng 1% số khối chứa công việc AI thật.

02 Nhu cầu trả tiền nằm ở suy luận, dữ liệu và quyền riêng tư

Bittensor khoảng 70 triệu USD doanh thu kiểm chứng được mỗi năm, Grass 17 triệu USD trong nửa đầu 2026, TEE đã vào vận hành; thanh toán của tác tử khoảng 5.000–11.000 USD mỗi tháng.

03 ASIC chỉ sinh lời khi quỹ thưởng lớn, luật lệ ổn định và có thị trường thứ hai

Trần lợi nhuận của nó là R × (1 − 1/k). Hiện chưa mạng nào hội đủ cả ba phân tích.

≈90 EH/s

Pearl ngày 13/11/2026: thông lượng số nguyên của khoảng 122.000 H100

≈1%

số khối Pearl chứa công việc AI thật (Blockworks)

4–11×

lượng phát hành của Bittensor so với doanh thu kiểm chứng được

AI × crypto qua sáu mảng: nhu cầu trả tiền thật nằm ở đâu

Chuỗi công việc hữu ích — 01

THẬT NHƯNG NHỎ

≈70 tr USD/năm

doanh thu bên ngoài kiểm chứng được của Bittensor

Tín hiệu: phát hành gấp 4–11× doanh thu; Pearl ≈1% khối AI

Cơ chế chạy được; tính hữu ích cần người trả tiền

Điện toán phi tập trung — 02

THẬT NHƯNG NHỎ

2,35 tr USD/năm

phí on-chain của Render, lớn nhất kiểm chứng được

Tín hiệu: vốn hóa ≈360× phí

Chủ yếu bán lại suy luận phổ thông

Huấn luyện và suy luận — 03

LÃI VỀ CỔ PHẦN

100+ tr USD/năm

Prime Intellect, công ty cổ phần dùng GPU tập trung

Tín hiệu: mainnet huấn luyện của Gensyn chưa chạy

Tiến bộ là thật; lợi nhuận về cổ phần

AI kiểm chứng được — 04

CÒN SỚM

398

người dùng trả tiền của đám mây TEE Phala, cuối 2025

Tín hiệu: chưa công bố zkML trả phí

TEE đã vào vận hành; zkML còn sớm

Thanh toán tác tử — 05

CHƯA CÓ NHU CẦU

5–11 nghìn USD/tháng

thanh toán x402 thực sự do tác tử khởi tạo

Tín hiệu: ≈200 tr giao dịch, phần lớn là tín hiệu

Hạ tầng có trước; nhu cầu thì chưa

Mạng dữ liệu — 06

CÓ NGƯỜI MUA

17 tr USD

doanh thu dữ liệu của Grass, nửa đầu 2026 (tự báo cáo)

Tín hiệu: vốn hóa ≈10× doanh thu năm hóa

Một trong số ít mảng có người mua thật

> Trợ cấp là nhiên liệu để khởi động, không phải mô hình kinh doanh. Crypto lo phân phối và thanh toán; cổ phần hưởng lợi nhuận, ít nhất là trong năm 2026.

Một phép nhân-cộng là một lần băm

Từ định lý đến mainnet

Komargodski và Weinstein (4/2025): bằng chứng công việc hữu ích cho mọi tích ma trận với chi phí phụ trội 1 + o(1). Mainnet Pearl chạy từ ngày 27/4/2026.

Cách hoạt động

Thợ đào nhân các ma trận có nhiễu, băm bộ tích lũy của mỗi ô thành một vé số, và chứng minh ô trúng bằng một bằng chứng Plonky2 dưới 60 KB.

Đơn vị

1 H là một phép nhân-cộng INT7 × INT7 → INT32 được ghi nhận: ngân sách bảo mật của blockchain được định giá bằng thông lượng tensor.

PEARL NGÀY 13/11/2026

89,77 EH/s

hashrate mạng (full node độc lập)

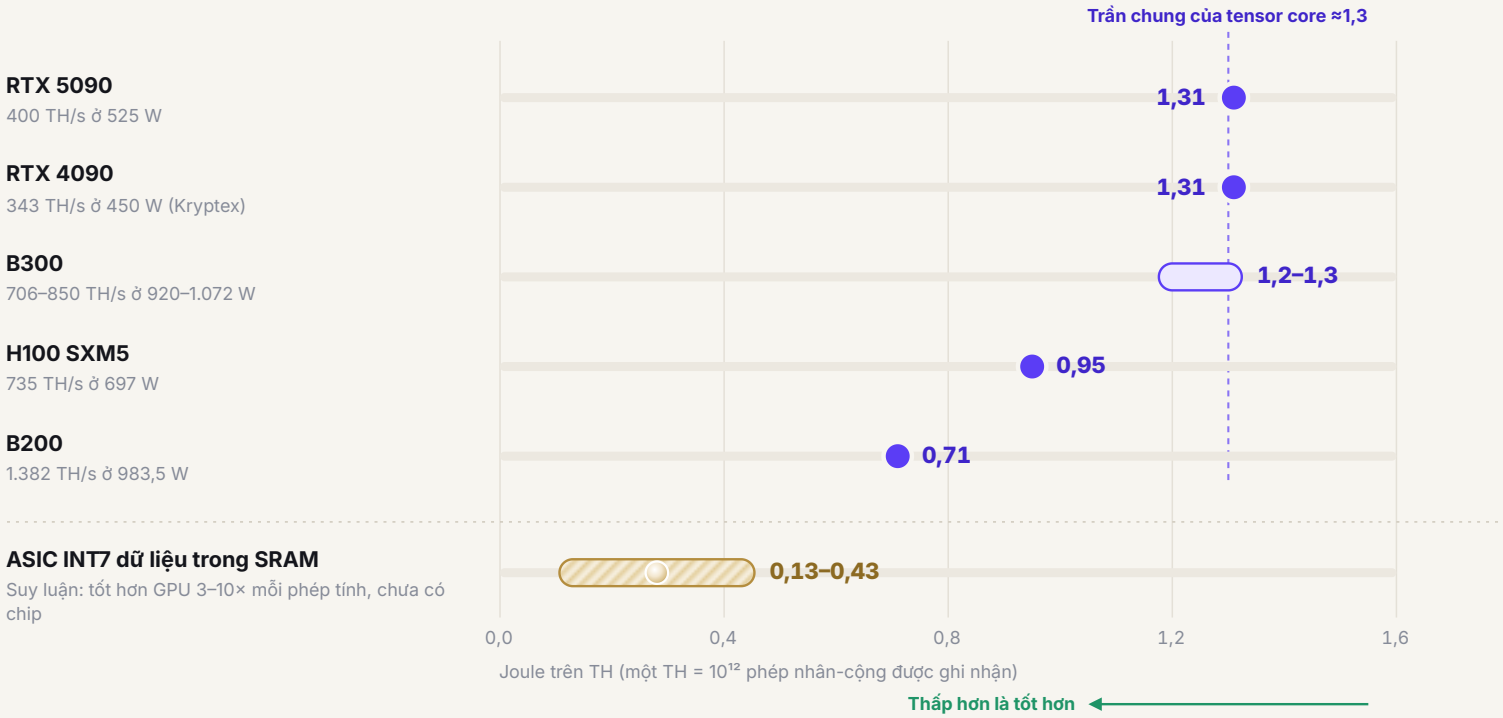
≈500 tr USD

vốn hóa; pha loãng hoàn toàn khoảng 2,8 tỷ USD

+78%

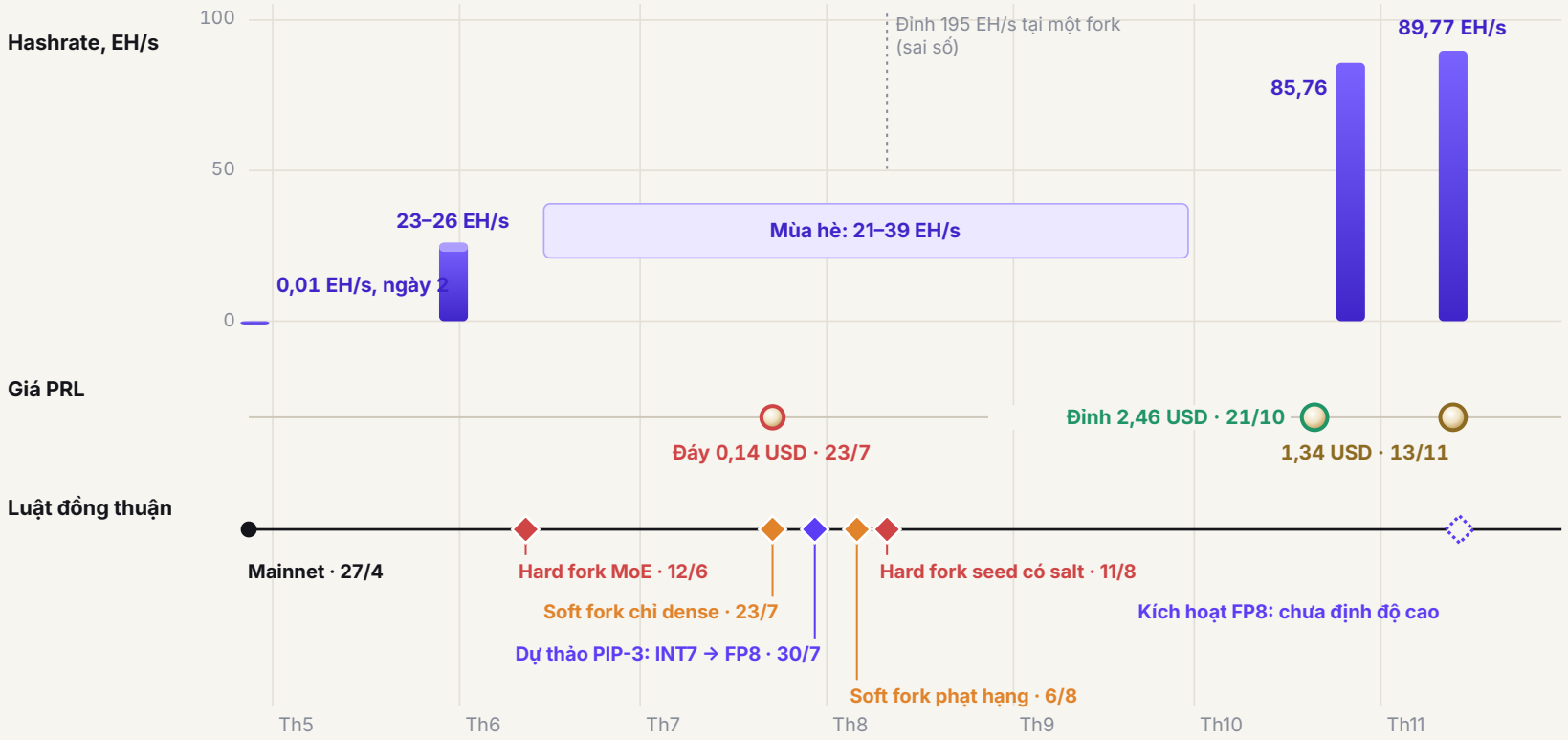
nguồn cung mới trong 12 tháng tới

Với công việc số nguyên của Pearl, các GPU chung một trần năng lượng



Card chơi game và B300 gặp nhau ở khoảng 1,3 J/TH; ngay cả B200 cũng chỉ đạt 0,71. Một con chip chỉ nhân số nguyên 7 bit từ SRAM có thể tốt hơn 3–10 lần phân tích.

Hashrate theo giá; luật lệ liên tục thay đổi



Khoảng 120 triệu đồng coin, 32% nguồn cung hiện nay, được đào trong sáu ngày đầu. Một RTX 5090 nay kiếm khoảng 5–6 USD mỗi ngày; thuê card chỉ để đào thì lỗ.

Định lý đứng vững; động lực thì không

CÔNG VIỆC HỮU ÍCH TỐN BAO NHIÊU

Phía suy luận: chi phí nhỏ

Pearl báo cáo thời gian tăng 5,08% cho Llama 70B trên bốn H200; một thử nghiệm độc lập đo được thông lượng giảm 9,2% trên một GPU Hopper.

CÔNG VIỆC HỮU ÍCH KIẾM ĐƯỢC BAO NHIÊU

Phía đào: quá ít hashrate

Card đang phục vụ suy luận chỉ đào được khoảng 27% so với máy đào chuyên dụng khi prefill, 5,4% với chat, khoảng 1% khi decode chiếm ưu thế.

Vì vậy thợ đào duy lý nhân các ma trận ngẫu nhiên. Tính hữu ích là bài toán thương mại ngoài chuỗi: phải có người trả tiền cho suy luận.

Bốn fork trong hai tháng, và tiếp theo là FP8

12/6
Hard fork mixture-of-experts (khối 71.935)

23/7
Soft fork chỉ dense (khối 91.630)

6/8
Soft fork phạt hạng: một số cấu hình mất 50–87,5% hashrate hiệu dụng

11/8
Hard fork seed có salt: máy đào cũ chỉ tạo share không hợp lệ

Chưa định
PIP-3: INT7 sang FP8, chỉ chấp nhận Hopper và Blackwell

2 trên 3

hữu ích cho AI, kiểm chứng được từng bit, trung lập phần cứng: một định dạng số chỉ đạt tối đa hai

13 bit

được giữ lại khi Hopper tích lũy FP8, có cắt bớt: không tái lập được giữa các thế hệ

≈82%

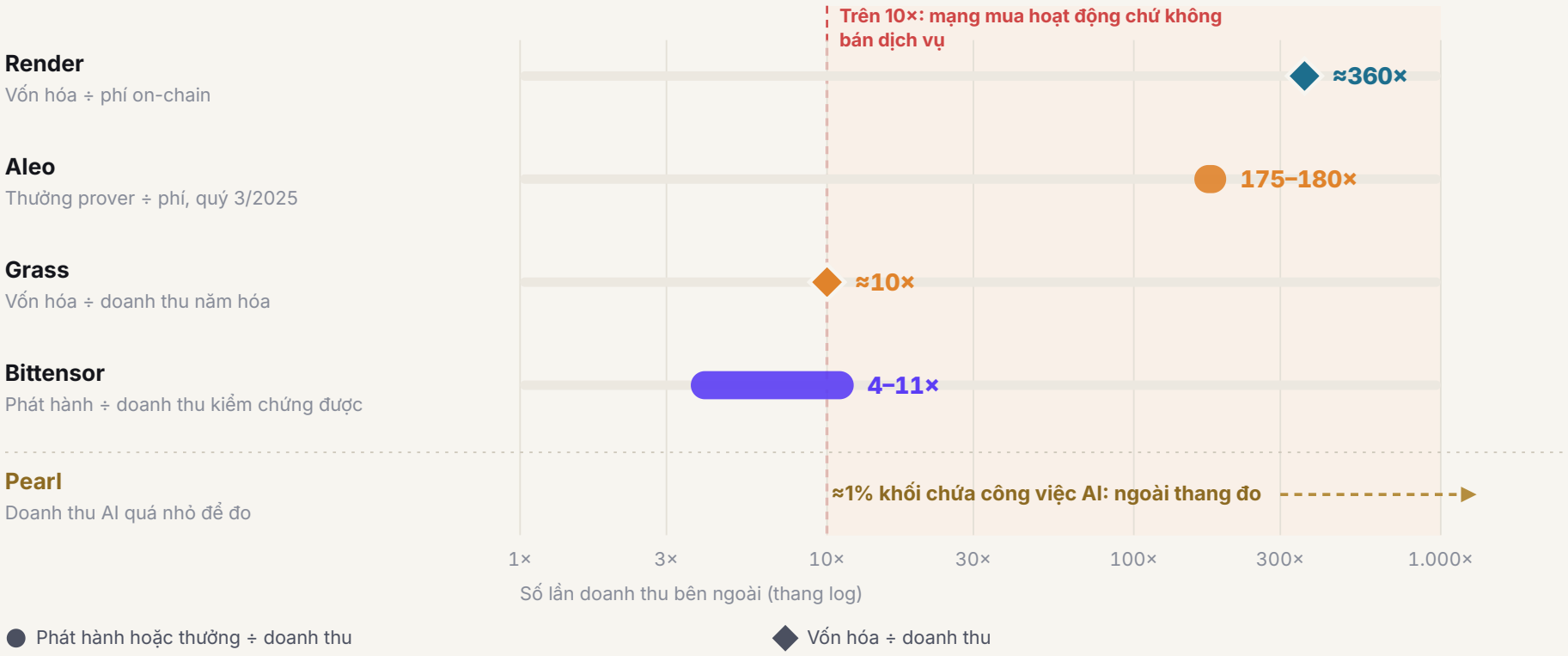
hashrate nằm trên card tiêu dùng mà PIP-3 sẽ không chấp nhận (mô hình độ tin cậy thấp)

Ai kiểm soát hàm công việc thì nắm số phận của nhà đầu tư phần cứng phân tích.

Cả ma trận lẫn zero knowledge đều không chặn được ASIC

Mạng	Điều đã xảy ra	Bài học
Kaspa	Tích ma trận 64 × 64 giữa các lần băm; ASIC xuất hiện sau 16 tháng, hiệu quả gấp khoảng 1.000× GPU; máy tốt nhất nay kiếm 0,86 USD mỗi ngày	Ma trận không hữu ích, cũng không chống được ASIC
Aleo	Câu đố SNARK tổng hợp; máy chuyên dụng sau năm tháng; thưởng prover gấp 175–180× phí	Công việc tổng hợp sống nhờ trợ cấp
Qubic	Tuyên bố huấn luyện AI; sức tính của nó đã đào Monero (tuyên bố 51%) và nay là Dogecoin	PoW đa dụng có thể bị chiếm từ bên ngoài
Bittensor	Chấm điểm theo trọng số stake; khoảng 70 triệu USD doanh thu kiểm chứng được mỗi năm so với hơn 300 triệu USD phát hành	Mạng duy nhất có doanh thu bên ngoài ở quy mô lớn
Monero, Sia, Ethereum	Fork biến ASIC thành phế liệu; sau The Merge chưa đến 20% thợ đào tìm được việc mới	Phần cứng có thị trường thứ hai thì sống sót

Bội số trợ cấp là phép kiểm tra sức khỏe trung thực nhất



> Trên khoảng 10 lần doanh thu bên ngoài, một mạng đang mua hoạt động chứ không bán dịch vụ. Doanh thu AI của Pearl quá nhỏ để đặt lên thang đo.

Nhu cầu thật nhưng nhỏ, và giá trị đang chuyển sang cổ phần

- **Thị trường điện toán**

  Phí on-chain của Render khoảng 2,35 triệu USD mỗi năm; H100 của Akash giá 2,73 USD/giờ rẻ hơn AWS và Google nhưng không rẻ hơn Vast.ai hay RunPod.

- **Huấn luyện**

  Huấn luyện 8–40 tỷ tham số qua internet; INTELLECT-3 đã quay về 512 H200 trên InfiniBand.

- **AI kiểm chứng được**

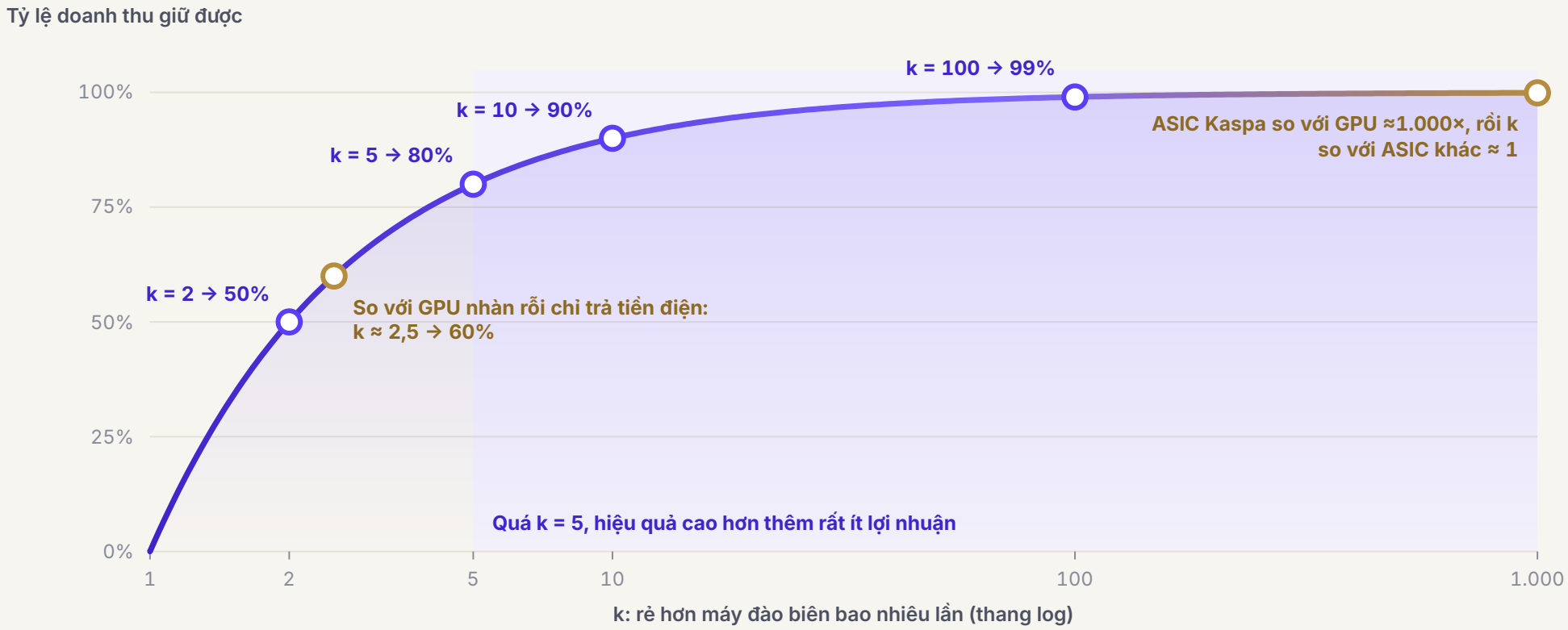
  TEE tốn thêm dưới 7% và đã vào vận hành; TEE.fail phá attestation bằng một interposer 1.000 USD; zkML chứng minh GPT-2 ở 174 token mỗi phút.

- **Tác tử và dữ liệu**

  Dòng x402 do tác tử khởi tạo khoảng 5.000–11.000 USD mỗi tháng; Grass có 17 triệu USD doanh thu dữ liệu trong nửa đầu 2026.

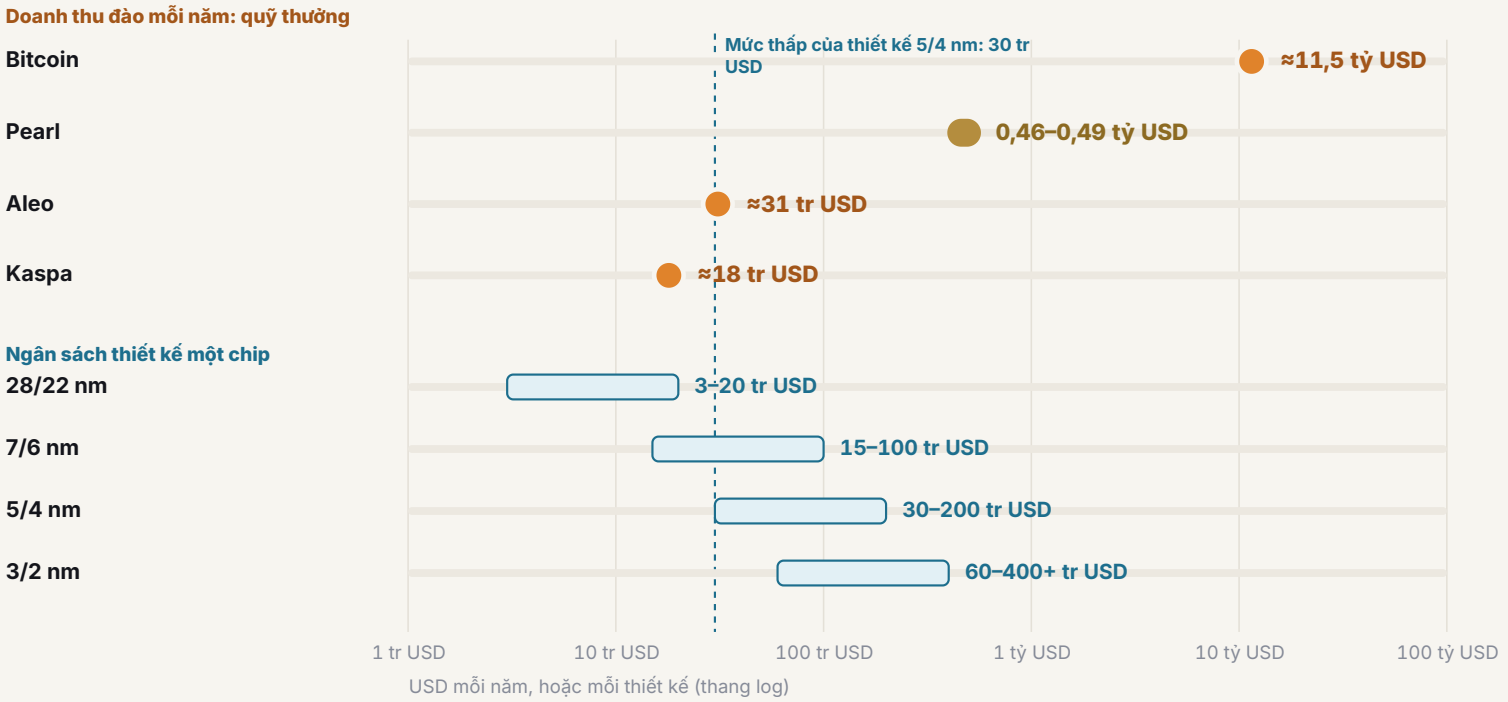
> **Bốn công ty cổ phần trên hạ tầng tập trung được định giá khoảng 8,5 tỷ USD, xấp xỉ toàn bộ token AI tầng giao thức và tầng ứng dụng.**

Vì sao hiệu quả thôi sinh lời sau k = 5



Trần lợi nhuận ≈ R × (1 − 1/k). Khi ASIC trở thành máy đào biên, k được tính lại so với các ASIC khác; so với GPU nhàn rỗi chỉ trả tiền điện, k khoảng 2,5 **ước tính**.

Phần lớn quỹ thưởng nhỏ hơn ngân sách thiết kế một con chip



Hashrate hoàn vốn h* = NRE ÷ [R × (1 − 1/k) × T × p]. Với Pearl, NRE 20 triệu USD và k = 5: khoảng 4,9 EH/s trong một năm nếu luật lệ giữ nguyên, 9,8 EH/s nếu xác suất giữ nguyên chỉ là năm ăn năm thua **ước tính**.

TỪNG LOẠI CHIP

Năm loại chip, năm câu trả lời

Chip đào PoUW kiểu Pearl

CHƯA NÊN

k: không HBM, không khối đa dụng: 3–10× mỗi phép tính (suy luận)

Thị trường thứ hai: suy luận AI, nếu là GEMM INT8/FP8 đa dụng

Đặt cược vào chip ưu tiên suy luận có thể đào, không phải chip INT7 chuyên dụng

Chip đào Bitcoin SHA-256

ĐỨNG NGOÀI

k: ≤1 so với các hãng đương nhiệm

Thị trường thứ hai: gần như không

Chỉ khi có điện riêng, dạng sản phẩm ngách, hoặc làm IP

Chip huấn luyện AI

IP, KHÔNG LÀM CHIP

k: 1,1–1,8× ở cùng độ chính xác

Thị trường thứ hai: —

Đóng gói và HBM quyết định; làm IP, chiplet, dịch vụ

Chip suy luận AI

NÊN LÀM

k: bỏ tính đa dụng: tuyên bố 10–20×

Thị trường thứ hai: đây là thị trường chính

Ký khách hàng trước, rồi mới chọn mức khóa chặt

Bộ tăng tốc chứng minh ZK

CHỜ VÀ PHÒNG HỘ

k: tuyên bố 10–100× GPU

Thị trường thứ hai: hạn chế ngoài ZK

Engine MSM và NTT lập trình được; FPGA trước

Bộ tăng tốc FHE

ĐÓN ĐẦU SỚM

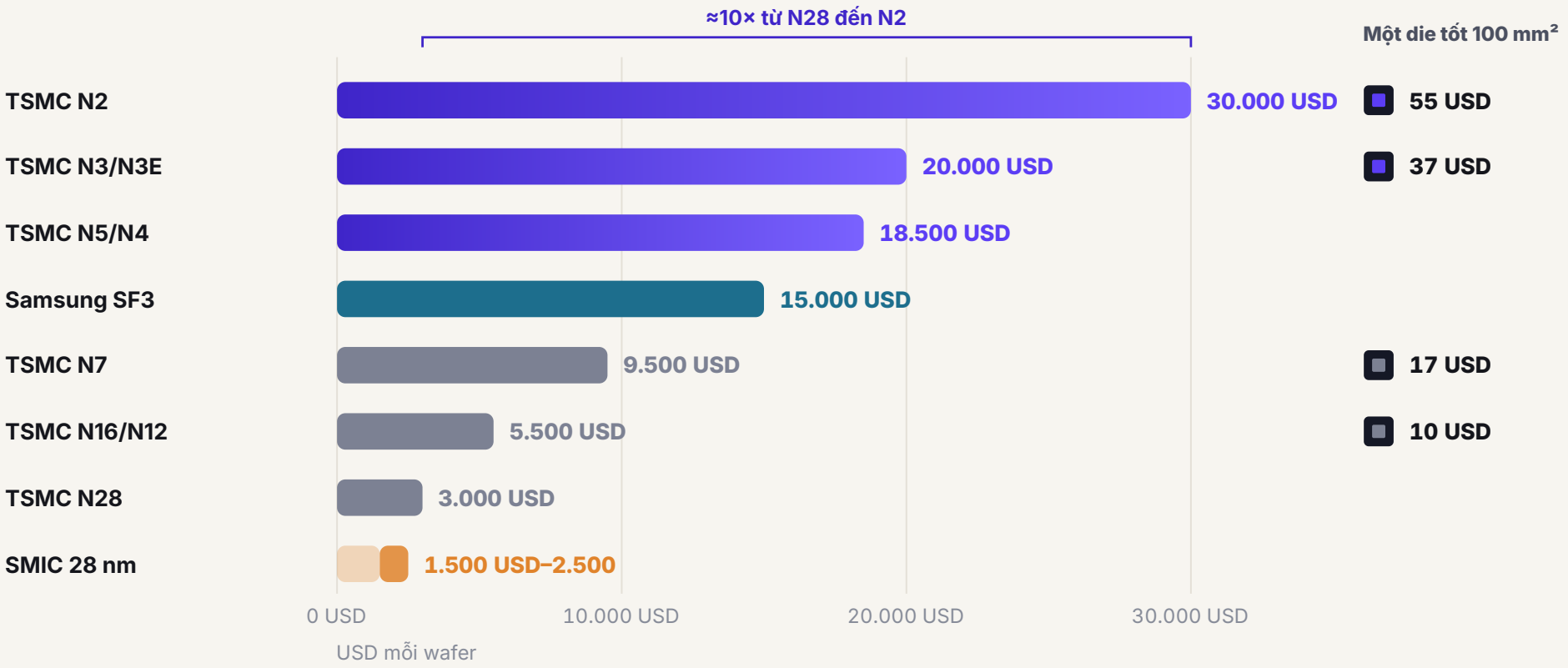
k: 1.074–5.547× một Xeon (Heracles)

Thị trường thứ hai: điện toán mã hóa, AI riêng tư

Cửa sổ 2027–2029; kiểm chứng với ít tài sản

> Hiện chưa mạng nào đồng thời có quỹ thưởng lớn, luật lệ ổn định và k trên 1: k của bitcoin thuộc về các hãng đương nhiệm, luật của Pearl kém ổn định nhất, quỹ thưởng của Kaspa và Aleo quá nhỏ.

Một wafer tiên tiến nhất đắt gấp mười wafer trưởng thành



> Hãy để chi phí vòng đời lớn nhất quyết định: năng lượng với chip tính toán chạy suốt ngày đêm, đóng gói và HBM với chip huấn luyện (63% chi phí một H100), lựa chọn SRAM hay HBM với chip suy luận.

TIẾN TRÌNH

Mỗi loại chip, mỗi đội ngũ một tiến trình

Chip	Đội nhỏ	Công ty quy mô lớn	Lộ trình Trung Quốc đại lục
Chip đào PoUW	MPW 28 hoặc 16/12 nm; sản lượng đầu N7/N6	N3/N3E hoặc Samsung SF3	SMIC N+1/N+2 dùng được, đắt hơn khoảng 40–50% mỗi die
Chip đào SHA-256	Đừng chạy theo tiến trình đầu bảng	Lớp 3 nm	Khó bắt kịp hiệu quả của tiến trình tiên tiến nhất
Chip huấn luyện AI	IP, chiplet, dịch vụ thiết kế	N3P rồi N2/A16 với CoWoS và HBM	SMIC N+2/N+3 với HBM nội địa
Chip suy luận AI	Lấy SRAM làm trung tâm ở N6/N7 hoặc N4/N5/SF4	N3/N2 với HBM, hoặc chiplet SRAM	SMIC N+2 với LPDDR hoặc GDDR
Bộ tăng tốc ZK và FHE	FPGA hoặc MPW 16/12 nm trước	N5/N3; FHE với HBM	ZK khả thi; FHE bị giới hạn bởi HBM

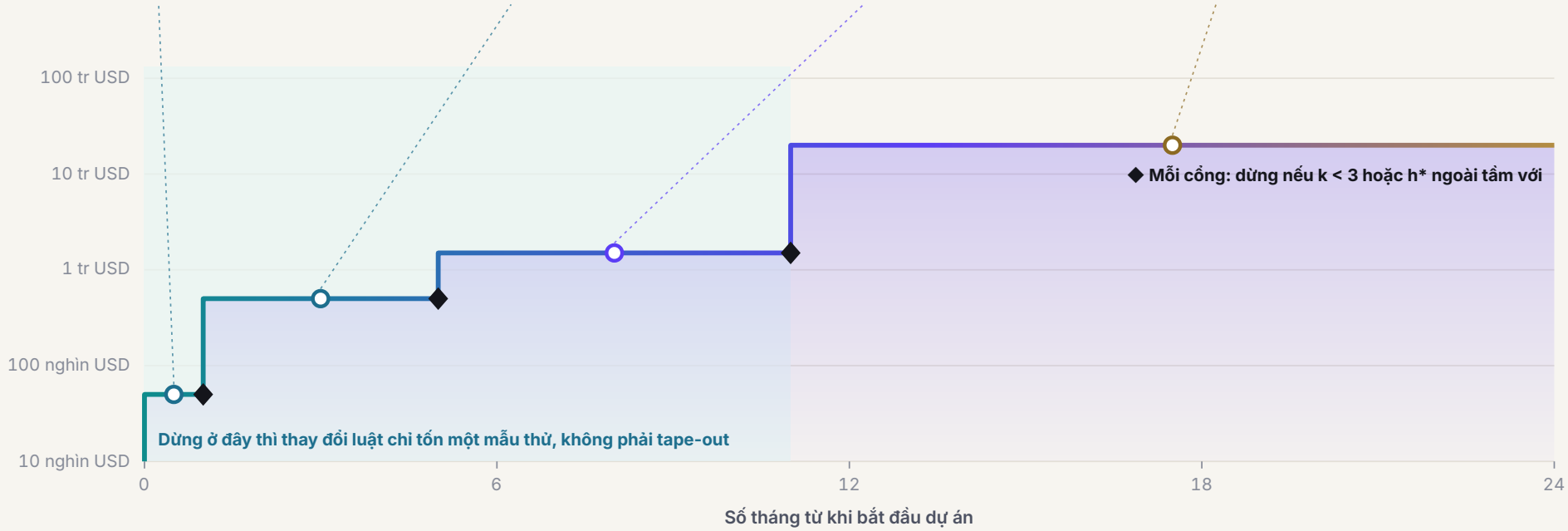
Mua bằng chứng trước khi mua mặt nạ

Cổng 0 · Mô hình
≈50 nghìn USD · vài tuần
J/op và $/op từ số học, SRAM và I/O

Cổng 1 · FPGA
≈0,5 tr USD · vài tháng
Tải thật từ đầu đến cuối: share hoặc token

Cổng 2 · MPW
≈1,5 tr USD · tháng 11
Silicon đo được thay cho mô hình

Cổng 3 · Mặt nạ và lô đầu
≈20 tr USD · tháng 24
Mặt nạ N7 ≈3,5 tr USD trong thiết kế 15–100 tr USD



Mô hình, FPGA và MPW đưa đến silicon đo được với khoảng 1,5 triệu USD trong mười một tháng; bộ mặt nạ và lô đầu tiên tốn gấp hơn mười lần. Dừng trước mặt nạ cho đến khi người mua hoặc phiên bản giao thức được cố định.

CƠ HỘI

Crypto là khách hàng đầu tiên, AI là thị trường thứ hai

Chip tùy biến

- Nhu cầu áp đảo đến từ AI: Broadcom khoảng 58 tỷ USD bán dẫn AI trong năm tài chính 2026
- Kỹ năng chip đào chuyển giao được: mảng điện áp thấp, cấp nguồn kilowatt, tape-out nhanh
- Thắng bằng kênh ra thị trường độc lập, không phải kênh đào coin

Phát triển AI

- Giờ công là hàng phổ thông: 25–49 USD mỗi giờ trên Clutch
- Giá trị nằm ở các khớp nối: kernel cho chip tùy biến, đào ngay trong dịch vụ suy luận
- Suy luận kiểm chứng được: TEE cộng kiểm tra ngẫu nhiên bằng ZK

Hệ thống AI × crypto

- Kiểm chứng phân tầng: TEE hoặc lấy mẫu, rồi tranh chấp hoặc kiểm tra ZK ngẫu nhiên
- Rào cản là phân phối và thanh khoản, không phải chi phí
- Ký với người mua sức tính trước khi làm token

TÔI SẼ LÀM GÌ

Ba hướng đi, mỗi hướng một bước đầu nhỏ

Hướng đi	Bước đầu
A. Bộ tăng tốc ưu tiên suy luận và cũng đào được	Nghiên cứu bốn tuần về J/op và $/op của một mảng INT8/FP8 dữ liệu trong SRAM so với RTX 5090 và H100, rồi đến FPGA hoặc MPW tiến trình trưởng thành
B. Dịch vụ silicon tùy biến: crypto trước, AI sau	Biến mô hình NRE và hòa vốn của báo cáo này thành các nghiên cứu khả thi có thu phí, có mốc và cố định phiên bản giao thức
C. Mạng AI-native có người mua trước khi có token	Một ghi chú thiết kế về kiểm chứng phân tầng và số học chuẩn trung lập với nhà cung cấp, viết cùng một người mua suy luận cụ thể

Tối đa ba hướng, mỗi hướng được thử trước khi cam kết tiền phân tích **.**

CẦN THEO DÕI

Điều cần theo dõi, và câu hỏi nên đặt ra

DẤU HIỆU

Bất cứ lúc nào	Độ cao kích hoạt FP8 của Pearl, và liệu chuỗi có tách đôi
Hằng tháng	Tỷ lệ khối Pearl chứa công việc AI
Hằng quý	Doanh thu kiểm chứng được của Bittensor so với phát hành; dòng x402 do tác tử khởi tạo
Tháng 1/2027	TSMC được cho là tăng giá 3–6% cho 2 nm và 3 nm
Quý 1/2027	FP4 trên Pearl, theo buổi AMA của các nhà sáng lập
Tháng 4/2028	Kỳ halving tiếp theo của bitcoin

CÂU HỎI NÊN ĐẶT RA

01 Ai kiểm soát hàm công việc?

02 Bội số trợ cấp là bao nhiêu?

03 Cổng tiếp theo tốn bao nhiêu, và điều gì khiến bạn dừng?

04 Hashrate hoàn vốn là bao nhiêu, khi tính cả p?

05 Thị trường thứ hai của chip là gì?

06 Khách hàng trả tiền đầu tiên là ai, tên cụ thể?

Đây là những lĩnh vực chúng ta có thể cùng khám phá, qua trao đổi kỹ thuật và hợp tác kỹ thuật: chip tùy biến, phát triển AI và hệ thống AI × crypto.

i@tanshuai.com · tanshuai.com