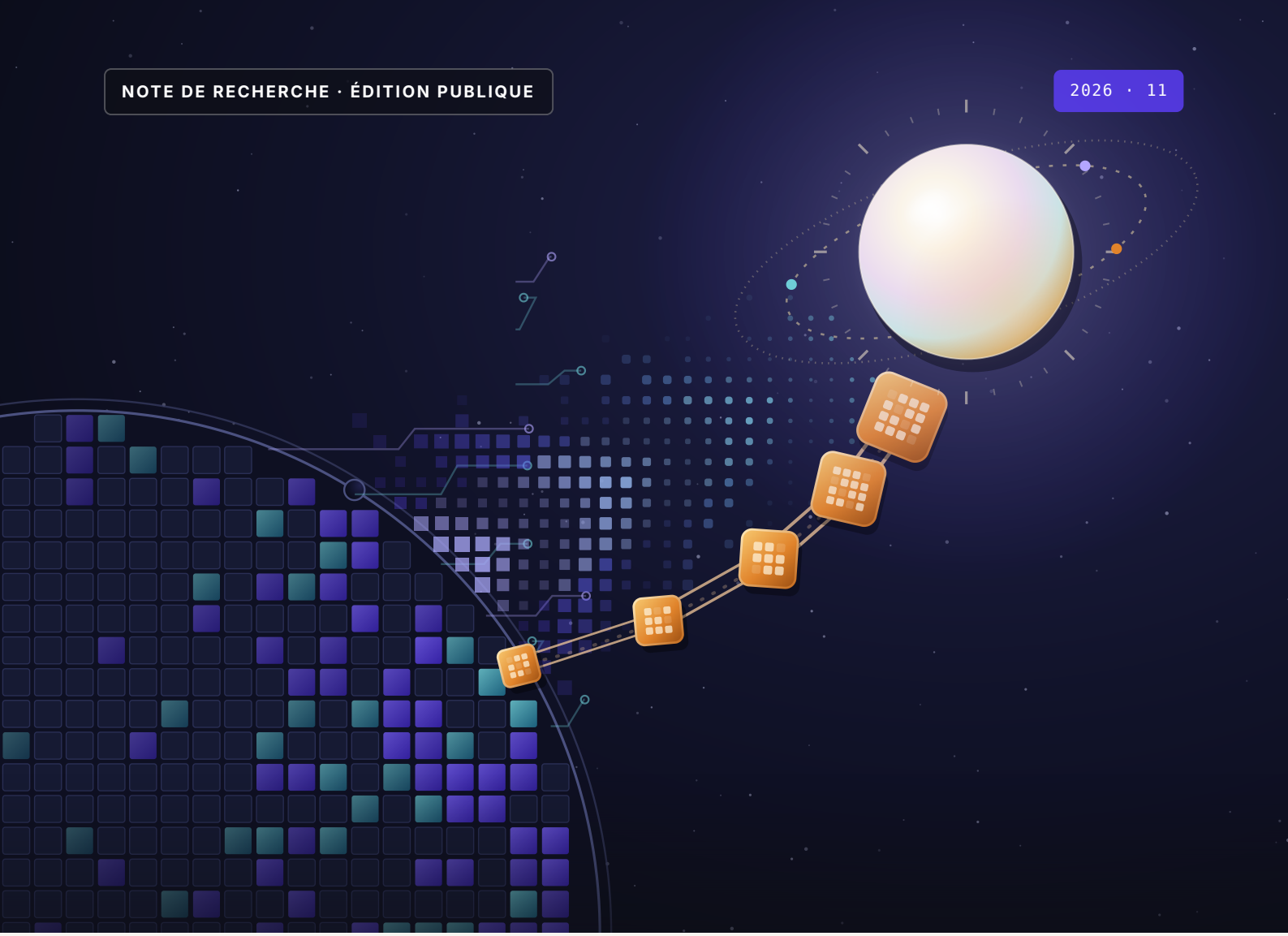


NOTE DE RECHERCHE · ÉDITION PUBLIQUE

2026 · 11

PREUVE DE TRAVAIL UTILE · CRYPTO × SILICIUM

Le calcul comme consensus

IA × crypto × puces en 2026, en partant de la preuve de travail utile de Pearl

Pearl · Travail utile · DePIN et agents · Économie des ASIC · Nœuds de gravure · Budgets par étapes

LE VERDICT

Pearl prouve que la multiplication matricielle peut sécuriser une blockchain, mais pas encore que le travail est utile. Dans l'IA × crypto, la demande payante se trouve dans l'inférence, les données et la confidentialité, et un ASIC n'est rentable que là où la cagnotte est grande, les règles stables et la puce dotée d'un second débouché.

Tan Shuai

Technologue et fondateur · Chercheur et développeur en IA

tanshuai.com · i@tanshuai.com

Données au 13 novembre 2026

Montants en dollars américains

SOMMAIRE

Sommaire



◆ **Synthèse** — Le théorème tient, l'incitation non — 04

01 **Périmètre et méthode** — Périmètre, méthode et quatre tests — 07

02 **Pearl · le mécanisme** — Pearl : la multiplication matricielle comme consensus — 08

03 **Pearl · l'économie** — Pearl : échelle, émission et écart d'utilité — 10

04 **Pearl · les règles** — Pearl : risque de règles, numérique et gouvernance — 12

05 **Chaînes à travail utile** — Chaînes à travail utile : ni les matrices ni la divulgation nulle de connaissance n'arrêtent les ASIC — 14

06 **Le paysage élargi** — Le reste de l'IA × crypto : une demande réelle mais faible, et une valeur qui passe aux actions — 17

07 **Économie des ASIC** — La question des ASIC : une formule, trois critères — 20

08 **Puce par puce** — Cinq types de puces, cinq réponses — 23

09 **Nœuds de gravure** — Nœuds de gravure : laisser le coût le plus lourd sur la durée de vie choisir — 27

10 **Budget par étapes** — Du modèle au silicium : un budget par étapes pour une première puce — 30

11 **Opportunités** — Où sont les opportunités, et ce que je ferais — 32

12 **Jalons** — Ce qu'il faut surveiller, de novembre 2026 à 2027 — 35

13 **Questions à poser** — Les questions que vous n'avez pas posées, mais auriez dû poser — 36

◆ **À propos de l'auteur** — Tan Shuai — 38

◆ **Sources** — 39

FIGURES

Figures



FIG. 01	L'IA × crypto en six secteurs : où la demande payante est réelle	06
FIG. 02	Sur le travail entier de Pearl, les GPU partagent un plafond énergétique	09
FIG. 03	Les six premiers mois de Pearl : le hashrate suit le prix, les règles changent sans cesse	11
FIG. 04	Le multiple de subvention : combien de fois ses revenus externes un réseau verse	16
FIG. 05	Pourquoi l'efficacité cesse de payer : un ASIC conserve au plus 1 − 1/k des revenus	21
FIG. 06	La plupart des cagnottes sont inférieures au budget de conception d'une seule puce	22
FIG. 07	Cinq types de puces, cinq réponses	26
FIG. 08	Un wafer de pointe coûte dix fois un wafer mature	28
FIG. 09	Acheter des preuves avant les masques : un budget par étapes pour une première puce	31

Étiquettes du texte

- **estimation** calcul ou déduction de l'auteur à partir de données citées
- **selon le fournisseur** publié par l'entreprise qui vend le produit, non vérifié de façon indépendante
- **analyse** interprétation des éléments par l'auteur

Ceci est une traduction de l'édition anglaise ; en cas de divergence, l'édition anglaise fait foi.

Données au 13 novembre 2026 · Montants en dollars américains

SYNTHÈSE

Le théorème tient, l'incitation non

LE VERDICT

Pearl montre que « le calcul comme consensus » fonctionne sur le plan technique, mais pas encore qu'il est utile.

≈90 EH/s
réseau de Pearl le 13 novembre 2026, où un H vaut une multiplication-accumulation INT7

≈1 %
des blocs de Pearl portent un vrai travail d'IA (Blockworks, 28 octobre 2026)

≈70 M$
par an de revenus externes vérifiables chez Bittensor, face à plus de 300 millions de dollars d'émissions

175–180×
les récompenses des prouveurs d'Aleo face à ses frais au troisième trimestre 2025

R × (1 – 1/k)
le plafond de profit d'un ASIC k fois moins coûteux que le mineur marginal
estimation

≈1,5 M$
pour obtenir du silicium mesuré en N7 via un modèle, un FPGA et un MPW, avant un jeu de masques à ≈3,5 M$
estimation

Pearl

Le 13 novembre 2026, le réseau de Pearl tournait à environ 90 EH/s. Dans son code de consensus, un « hash » est une multiplication-accumulation entière, si bien que le réseau équivaut au débit entier d'environ 122 000 H100. Pourtant, seuls environ 1 % des blocs portent un vrai travail d'IA, et l'unique partenariat d'inférence public n'est plus proposé sur l'API serverless de son fournisseur. La cause est économique, non cryptographique : un GPU qui mine tout en servant de l'inférence n'obtient qu'environ 1 % à 27 % du hashrate d'un mineur dédié, si bien que les mineurs rationnels multiplient des matrices aléatoires.

Demande payante

Dans l'ensemble IA × crypto, la demande payante vérifiable apparaît à trois endroits : vendre de l'inférence, vendre des données et vendre de la confidentialité. Les revenus externes vérifiables de Bittensor atteignent environ 70 millions de dollars par an, dont environ quatre cinquièmes proviennent de sous-réseaux de calcul et d'infrastructure. Grass déclare 17 millions de dollars de revenus de données au premier semestre 2026. L'inférence confidentielle dans des environnements d'exécution de confiance est en production. À l'inverse, les paiements réellement initiés par des agents d'IA représentent environ 5 000–11 000 dollars par mois, et ni le zkML ni l'entraînement décentralisé n'ont encore de demande payante déclarée.

Subvention et valeur

Le multiple de subvention est le bilan de santé le plus honnête. Les récompenses des prouveurs d'Aleo représentaient environ 175–180 fois ses frais ; les émissions de Bittensor valent environ 4–11 fois ses revenus vérifiables ; la valeur de marché de Render atteint environ 360 fois ses frais on-chain ; et les quelque 460 millions de dollars annuels de revenus des mineurs de Pearl sont pour l'essentiel de la subvention de bloc, presque sans lien avec une demande d'IA payante. La subvention est un carburant pour un démarrage à froid, pas un modèle écono-

mique. Pendant ce temps, la valeur passe des tokens aux actions : quatre sociétés par actions qui utilisent une infrastructure centralisée et une distribution native crypto sont valorisées à environ 8,5 milliards de dollars, à peu près autant que l'ensemble des tokens d'IA des couches protocole et application.

ASIC

Pour les puces, le plafond de profit d'un ASIC est à peu près le revenu par unité de calcul multiplié par (1 − 1/k), où k indique combien de fois il est moins coûteux que le matériel marginal. Ce terme étant concave, l'efficacité au-delà de k = 5–10 n'ajoute presque aucun profit ; ce qui décide de l'issue, c'est la taille de la cagnotte, la durée de vie des règles et l'existence d'un second débouché pour la puce. Mes réponses par type de puce : ne pas lancer maintenant le tape-out d'une puce Pearl INT7 dédiée, et parier plutôt sur un accélérateur conçu d'abord pour l'inférence et capable de miner ; rester à l'écart, en tant que nouvel entrant, du marché ouvert des machines de minage bitcoin ; ne pas concevoir de puce d'entraînement autonome en petite équipe ; concevoir des puces d'inférence lorsqu'un client est signé ; se couvrir sur les accélérateurs ZK et se positionner tôt sur le FHE.

Nœuds et budgets

Le nœud de gravure doit être choisi d'après le poste le plus lourd du coût de la puce sur sa durée de vie : l'énergie pour les puces qui tournent en continu, le packaging et la mémoire pour les puces d'entraînement, le choix entre SRAM et HBM pour les puces d'inférence. Et une première puce doit s'acheter par étapes, un modèle, un FPGA, un wafer multiprojet et seulement ensuite un jeu de masques, afin qu'un changement du protocole coûte le prix d'un prototype et non celui d'un tape-out. Les opportunités se trouvent au croisement des domaines : la crypto comme premier client, l'IA comme second débouché.

Question	Jugement	Chiffres clés (sources dans les chapitres)
① **Pearl fonctionne-t-il, et son travail est-il utile ?**	Il fonctionne : une blockchain sécurisée à grande échelle par la multiplication matricielle. Pas encore utile : presque tout le minage porte sur des matrices aléatoires	Environ 90 EH/s le 13 novembre 2026 ; 1 H = 1 multiplication-accumulation INT7 ; environ 1 % des blocs avec du travail d'IA
② **Où est la vraie demande payante en IA × crypto ?**	Inférence, données et confidentialité. Paiements d'agents, zkML et entraînement décentralisé en sont encore à l'avant-demande	Bittensor environ 70 millions de dollars par an de revenus vérifiables ; Grass 17 millions de dollars au S1 2026 ; paiements d'agents environ 5 000–11 000 dollars par mois
③ **Un ASIC vaut-il la peine d'être conçu ?**	Seulement là où la cagnotte est grande, les règles stables et la puce dotée d'un second débouché. Aujourd'hui, aucun réseau n'offre les trois	Plafond de profit ≈ R × (1 − 1/k) ; bitcoin environ 11,5 milliards de dollars par an de revenus des mineurs, Pearl environ 0,46–0,49 milliard, Kaspa environ 18 millions
④ **Quelles puces une petite équipe doit-elle concevoir ?**	Des accélérateurs conçus d'abord pour l'inférence et capables de miner ; des puces d'inférence pour un client signé ; des travaux précoces sur le FHE. Pas de machines de minage bitcoin pour le marché ouvert ni de puces d'entraînement	Le meilleur ASIC Kaspa rapporte 0,86 dollar par jour ; budgets de conception en 3/2 nm de 60–400 millions de dollars ou plus
⑤ **Quel nœud de gravure ?**	Le coût le plus lourd sur la durée de vie décide : énergie, packaging et mémoire, ou architecture mémoire	TSMC N2 environ 30 000 dollars le wafer contre environ 3 000 pour N28 ; HBM et CoWoS environ 63 % du coût d'un H100
⑥ **Que coûte une première puce, et comment l'acheter ?**	En quatre étapes : un modèle, un FPGA, un wafer multiprojet, puis les masques. S'arrêter avant les masques si les règles ou l'acheteur ne sont pas fixés	Une place MPW commerciale 50 000–500 000 dollars ; masques N7 environ 3,5 millions de dollars ; un budget de conception en 7/6 nm de 15–100 millions de dollars

FIG. 01 **L'IA × crypto en six secteurs : où la demande payante est réelle**

Demande payante vérifiable face aux signaux de subvention et de valorisation, novembre 2026

Chaînes à travail utile 01

RÉEL MAIS FAIBLE

≈70 M$ par an

revenus externes vérifiables de Bittensor

Signal : émissions 4–11× les revenus ; Pearl ≈1 % de blocs IA

Les mécanismes marchent ; l'utilité veut des payeurs

Calcul décentralisé 02

RÉEL MAIS FAIBLE

2,35 M$ par an

frais on-chain de Render, les plus élevés vérifiables

Signal : valeur de marché ≈360× les frais

Surtout de la revente d'inférence banalisée

Entraînement et inférence 03

PROFIT EN ACTIONS

100 M$+ par an

Prime Intellect, société par actions sur GPU centralisés

Signal : mainnet d'entraînement de Gensyn non lancé

Progrès réels ; le profit va aux actions

IA vérifiable 04

PRÉCOCE

398

utilisateurs payants du cloud TEE de Phala, fin 2025

Signal : aucun zkML payant déclaré

TEE en production ; zkML précoce

Paiements d'agents 05

AVANT LA DEMANDE

5–11 k$ par mois

paiements x402 réellement initiés par des agents

Signal : ≈200 M de transactions, surtout de la signalisation

Les rails d'abord ; la demande n'a pas suivi

Réseaux de données 06

VRAIS ACHETEURS

17 M$

revenus de données de Grass, S1 2026 (autodéclarés)

Signal : valeur de marché ≈10× les revenus annualisés

L'un des rares secteurs à vrais acheteurs

Source: DefiLlama, SubConnect (via hoge.gg), TRM Labs, Phala, Dealroom, SiliconANGLE, CoinGecko, Blockworks ; synthèse de l'auteur.

CHAPITRE 01

01 Périmètre, méthode et quatre tests

PÉRIMÈTRE ET MÉTHODE

Pourquoi ce rapport, et pourquoi maintenant

En avril 2026, une nouvelle blockchain, Pearl, a mis en production une idée ancienne : faire concourir les mineurs pour les blocs en effectuant les multiplications matricielles sur lesquelles repose l'IA, au lieu de calculer des hashs. En six mois, son réseau a atteint environ 90 EH/s et son token une valeur de marché d'environ 500 millions de dollars. Dans le même temps, les places de marché de calcul pour l'IA, les réseaux d'entraînement décentralisé, les infrastructures de paiement pour agents et les réseaux de données ont levé des milliards, et les mineurs de bitcoin ont commencé à convertir leurs sites électriques en centres de données d'IA. Je voulais savoir lesquels ont de vrais clients, si le silicium sur mesure a un sens quelque part dans cet ensemble, et quel nœud de gravure et quel fondeur une équipe de conception de puces peut réellement utiliser.

Quatre tests pour tout projet IA × crypto

Une étude de sécurité de Gridcoin publiée en 2017 soutenait qu'un calcul utile ne peut sécuriser une blockchain que s'il est difficile à résoudre, facile à vérifier et facile à récompenser. J'ajoute un quatrième test. Ensemble, ils déterminent l'essentiel de ce qui suit.

01 **Difficile à résoudre.** Le calcul honnête doit être le moyen le moins coûteux de produire une preuve valide ; il ne doit exister aucun raccourci.

02 **Facile à vérifier.** Vérifier le travail doit coûter bien moins cher que de l'effectuer, sinon le réseau paie deux fois.

03 **Facile à récompenser.** Le protocole doit pouvoir calculer la récompense à partir de la seule preuve, sans faire confiance à un tiers.

04 **Quelqu'un paie pour le résultat.** Si aucun client extérieur au réseau ne paie pour le résultat, le travail « utile » n'est qu'une subvention assortie d'un récit.

Les puzzles de hachage réussissent les trois premiers tests et échouent au quatrième. Les vraies tâches d'IA réussissent le quatrième mais échouent au deuxième : prouver un calcul d'IA à divulgation nulle de connaissance coûte de 100 à 10 000 fois le calcul lui-même, et il a été démontré que les schémas de preuve d'apprentissage peuvent être falsifiés à peu de frais. La multiplication matricielle bruitée de Pearl est la première passerelle crédible qui réussit le test de vérification tout en conservant la forme du calcul d'IA.

Ma méthode

Il s'agit d'une recherche documentaire, menée comme j'examinerais une puce ou un plan d'affaires. J'ai lu le code de Pearl, ses propositions de protocole et son livre blanc ; les données de la chaîne issues d'un nœud complet indépendant ; les documents déposés par les entreprises et leurs communiqués de résultats ; des articles de cryptographie, d'architecture des ordinateurs et d'apprentissage automatique ; les données de prix des fondeurs et des fournisseurs ; ainsi que la presse spécialisée et technologique. Chaque chiffre porte sa date. Lorsque je cal-

cule moi-même un chiffre, je le dis et le marque estimation. Les chiffres que seul le vendeur communique sont marqués selon le fournisseur ; mon interprétation des éléments est marquée analyse. Les montants sont en dollars américains. La situation est celle du 13 novembre 2026.

CHAPITRE 02

02

PEARL · LE MÉCANISME

Pearl : la multiplication matricielle comme consensus

L'idée, du théorème au mainnet

En avril 2025, Ilan Komargodski et Omri Weinstein ont publié « Proofs of Useful Work from Arbitrary Matrix Multiplication ». Ils construisent une preuve de travail utile pour tout produit matriciel avec un surcoût de seulement 1 + o(1) fois une multiplication naïve ; sa sécurité repose sur la difficulté à résoudre des lots d'équations linéaires aléatoires de faible rang, ce qui est une conjecture et non une démonstration. L'article nomme la vraie difficulté du domaine : une preuve de travail utile avec un surcoût d'un facteur constant est triviale mais inutile ; le plus dur est de faire en sorte que le calcul honnête reste le moyen le moins coûteux de gagner lorsque les mineurs choisissent leurs propres entrées.

Le mainnet de Pearl a démarré le 27 avril 2026 à 09:00 UTC. Les mineurs calculent un produit bruité (A + E)(B + F), où le bruit de faible rang est dérivé d'un engagement BLAKE3, puis retirent le bruit pour retrouver A·B. L'état de l'accumulateur de chaque tuile de sortie est haché en un ticket de loterie ; une tuile gagnante est ouverte et prouvée par une preuve à divulgation nulle de connaissance Plonky2 récursive de moins de 60 Ko, à l'intérieur d'un certificat de bloc plafonné à 65 Ko. La vérification porte sur une petite preuve de taille constante, indépendante de la taille des matrices : aussi peu coûteuse à vérifier qu'un puzzle de hachage, alors que le calcul lui-même peut être utile.

Une multiplication-accumulation vaut un « hash »

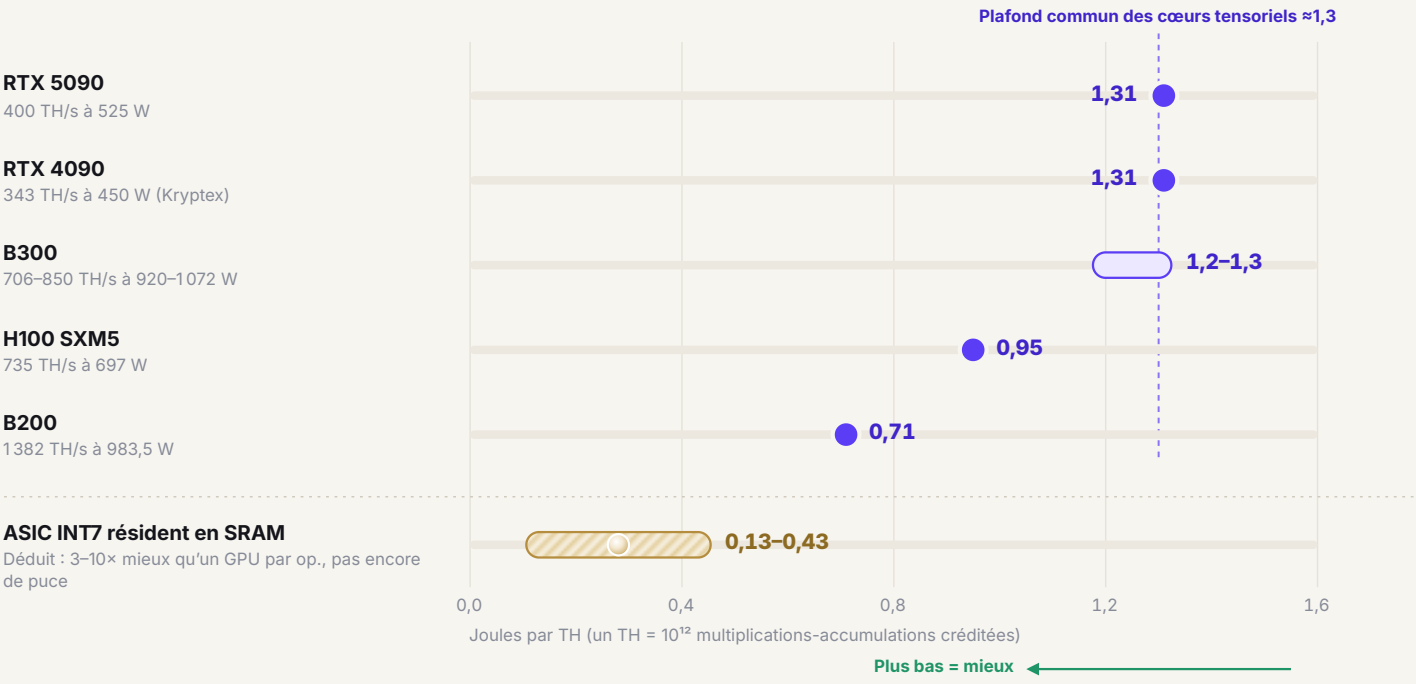
Pour un concepteur de puces, le fait le plus important est l'unité. Le code de consensus de Pearl met à l'échelle la cible de chaque tentative par la taille de la tuile multipliée par la longueur du produit scalaire, si bien que **1 H correspond exactement à une multiplication-accumulation (MAC) INT7 × INT7 → INT32 créditée**. Depuis le bloc 96 251, une pénalité de rang impose un rang de bruit d'au moins 128 ; au rang minimal, 1 H vaut toujours 1 MAC. Les « EH/s » de Pearl signifient donc littéralement 10^18 multiplications-accumulations entières par seconde : **pour la première fois, le budget de sécurité d'une blockchain est exprimé dans le débit tensoriel des accélérateurs d'IA.** Les hashrates de GPU mesurés se situent entre 60 % et 95 % du pic de multiplications-accumulations INT8 denses de chaque carte, ce qui concorde avec cette lecture.

Matériel	Hashrate Pearl	Puissance	Joules par billion de MAC (J/TH)	Source (date)
RTX 5090	400 TH/s	525 W	1,31 estimation	Kryptex (12 novembre)
RTX 4090	environ 305 TH/s (fourchette communautaire 280–330) ; Kryptex indique 343 TH/s	450 W (Kryptex)	1,31 (chiffres Kryptex) estimation	prlstats (3 octobre) ; Kryptex
H100 SXM5	735 TH/s	697 W	0,95 estimation	PeakMiner, via prlstats (19 septembre)

Matériel	Hashrate Pearl	Puissance	Joules par billion de MAC (J/TH)	Source (date)
B200	1,382 PH/s	983,5 W	0,71 estimation	PeakMiner, via prlstats
B300	706–850 TH/s	920–1 072 W	1,2–1,3 estimation	prlstats ; SimeonOnSecurity

FIG. 02 **Sur le travail entier de Pearl, les GPU partagent un plafond énergétique**

Joules par billion de multiplications-accumulations INT7 (J/TH) au hashrate Pearl mesuré ; plus bas = mieux



Source: Kryptex (12 novembre 2026) ; benchmarks PeakMiner et SRBMiner via prlstats ; SimeonOnSecurity ; la fourchette ASIC est une déduction de l'auteur, pas une mesure.

Deux détails de ce tableau comptent pour les concepteurs de puces. D'abord, les cartes grand public RTX 4090 et 5090 et le B300 pour centres de données convergent vers environ 1,3 J/TH, et même la meilleure carte, le B200, n'atteint qu'environ 0,71 J/TH. Un GPU de centre de données coûtant des dizaines de milliers de dollars dépense presque autant d'énergie par multiplication-accumulation entière qu'une carte de jeu, ce qui indique un plafond commun de l'énergie des cœurs tensoriels par MAC INT8, et c'est précisément ce plafond qu'une puce dédiée pourrait percer. Ensuite, le B300 est plus lent que le B200 : NVIDIA a fortement réduit le débit INT8 dans Blackwell Ultra, si bien que le travail entier doit emprunter un chemin plus lent sur cette puce. Le silicium d'IA dominant s'éloigne de l'INT8 au profit du FP8 et du FP4. Quand le format numérique choisi par un consensus diverge de la direction prise par le matériel, le protocole est contraint de changer ses règles, et c'est la racine du fork FP8 comme de mon jugement sur les ASIC plus loin dans ce rapport.

CHAPITRE 03

03 PEARL · L'ÉCONOMIE
Pearl : échelle, émission et écart d'utilité

De 0,01 à 90 EH/s en six mois

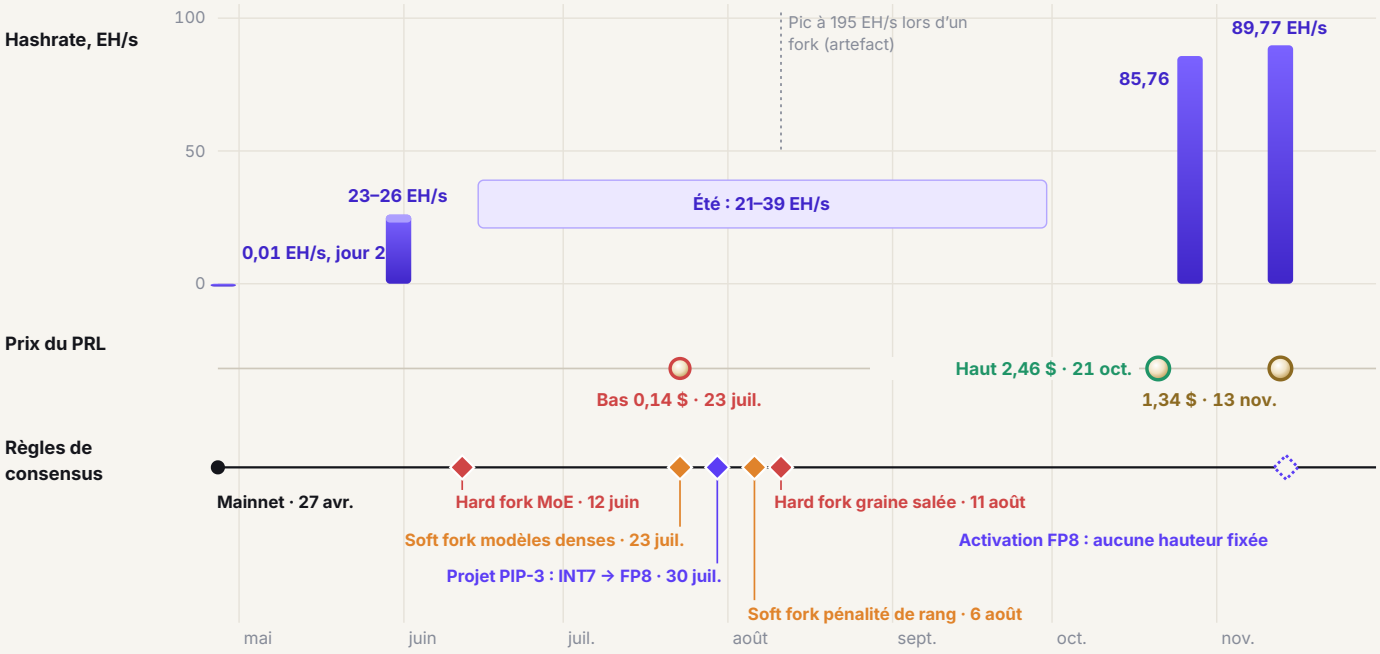
Indicateur	Valeur (13 novembre 2026)	Base et source
Hashrate du réseau	Environ 90 EH/s (référence 89,77 ; fourchette du jour 88,20–93,22 ; pic intrajournalier du 12 novembre de 110–123)	Nœud complet indépendant prlstats ; Kryptex
Difficulté et hauteur de bloc	58 841 982 et 142 530	prlstats
Matériel équivalent	Environ 122 000 H100 ou 225 000 RTX 5090	À 735 et 400 TH/s par carte estimation
Prix, valeur de marché et valeur entièrement diluée	1,32–1,34 dollar ; environ 500 millions de dollars ; environ 2,8 milliards de dollars (rang CoinGecko 101)	CoinGecko
Plus haut et plus bas historiques	2,46 dollars (21 octobre) et 0,1399 dollar (23 juillet)	CoinGecko
Pièces minées	377,6 millions de PRL, 17,98 % du plafond de 2,1 milliards	prlstats
Subvention de bloc et émission quotidienne	Environ 2 173 PRL et environ 0,97–1,03 million de PRL	prlstats ; Kryptex
Nouvelle offre sur les 12 prochains mois	Environ 78 % de l'offre actuelle	prlstats
Revenus des mineurs	Environ 1,25 million de dollars par jour (modèle), environ 460 millions de dollars par an	prlstats
Revenu par unité de hashrate, R	Environ 14–15 dollars par PH/s et par jour ; une RTX 5090 rapporte environ 5–6 dollars par jour	estimation ; Kryptex
Blocs portant un vrai travail d'IA	Environ 1 % (estimation du 28 octobre)	Blockworks
Concentration des pools	Kryptex a trouvé 46,4 % des blocs des dernières 24 heures ; les deux premiers pools environ 62 %	prlstats
Concentration des détenteurs	Coefficient de Gini des adresses 0,969 ; les 100 premières adresses détiennent 53,0 %	prlstats

Le hashrate de Pearl est presque une fonction retardée de son prix. Le lendemain du lancement, le réseau tournait à environ 0,01 EH/s ; à la fin mai, il atteignait 23–26 EH/s. Pendant l'été, le prix a chuté d'environ 86 % par rapport à son sommet de mai et le hashrate a stagné à 21–39 EH/s. Après une hausse du prix d'environ dix fois en septembre et en octobre, le hashrate a atteint 85,76 EH/s le 27 octobre et environ 90 EH/s en novembre. Un pic à 195 EH/s le 11 août a coïncidé avec un fork et doit être lu comme un artefact d'estimation. La consommation

d'énergie a suivi : une étude l'estimait à environ 112 MW pour l'ensemble du réseau à 24 EH/s, et un modèle de prlstats la situe aujourd'hui à environ 190–215 MW.

FIG. 03 **Les six premiers mois de Pearl : le hashrate suit le prix, les règles changent sans cesse**

Hashrate du réseau, trois repères de prix et chaque changement de consensus, du 27 avril au 13 novembre 2026



Source: Nœud complet et séries prlstats ; Kryptex ; CoinGecko ; params.go et PIP-3 de Pearl. Le pic du 11 août est un artefact d'estimation lors d'un fork.

L'émission amplifie la réflexivité

L'offre est plafonnée à 2,1 milliards de pièces et il n'y a pas de halving : la subvention de bloc décroît de façon continue selon E(t) = S·H / ((t + H)(t + H − 1)) avec H = 650 226. Mais l'émission se compte en blocs, non en temps, et la difficulté initiale était trop basse et s'est ajustée lentement. **Environ 44,06 millions de pièces ont été minées le premier jour et environ 120 millions au cours des six premiers jours**, soit environ 32 % de l'offre actuelle estimation. Les blocs récents, à environ 171–183 secondes, arrivent encore plus vite que la cible de 194 secondes. Résultat : l'offre augmentera d'environ 78 % au cours des 12 prochains mois, une dilution que mineurs, fabricants de matériel et détenteurs doivent tous affronter.

L'économie des mineurs en découle. Les revenus des mineurs à l'échelle du réseau sont d'environ 1,25 million de dollars par jour (modèle) ; l'émission quotidienne de Kryptex multipliée par le prix actuel donne environ 1,35 million de dollars par jour, d'où **un revenu par unité de hashrate R ≈ 14–15 dollars par PH/s et par jour** estimation. Une RTX 5090 rapporte environ 5–6 dollars par jour. Un H100 ou un H200 rapporte environ 11 dollars, alors que la location publique à la demande la moins chère coûte environ 2,09 dollars par carte et par heure, soit environ 50 dollars par jour : louer des cartes uniquement pour miner fait perdre beaucoup d'argent. R fluctue fortement : selon la série de prlstats, il était d'environ 10 dollars en août et en septembre et brièvement d'environ 33 dollars à la mi-octobre estimation. Dans les semaines suivant le lancement, le revenu quotidien estimé d'une RTX 5090 a été divisé par deux, passant d'environ 33,80 dollars à environ 17,19 dollars. Louer des GPU dans le cloud pour miner n'a été rentable que durant les premières semaines et pendant la flambée des prix d'octobre.

Le théorème tient, l'incitation non

Le vrai problème est la demande, non la cryptographie. La seule intégration commerciale publique de Pearl est un point d'accès Gemma-4-31B-it-pearl de Together AI, lancé le 15 mai à un prix inférieur de plus de 25 % au tarif habituel, la remise étant financée par la valeur future attendue de l'émission. Le 13 novembre, la page du modèle indiquait qu'il **n'est pas disponible sur l'API serverless de Together**. En juin, un recensement de 8 012 instances de minage en ligne, soit environ 21 % du réseau, a conclu que le réseau ne produisait aucun calcul d'IA utile (l'auteur qualifie les éléments de suggestifs, non concluants), et a montré que des parts construites à partir de matrices purement aléatoires étaient acceptées par les pools. Le 28 octobre, Blockworks estimait que **seuls environ 1 % des blocs portent un vrai travail d'IA** ; sans lui, écrivait-il, Pearl n'est qu'un bitcoin plus petit, moins sûr et à l'inflation plus rapide.

La raison se calcule. Le surcoût côté inférence est réellement faible : Pearl indique une hausse de 5,08 % du temps de bout en bout pour Llama 70B sur quatre H200, et un test indépendant a mesuré une perte de débit de 9,2 % sur un GPU Hopper. Le côté minage rapporte trop peu. Le plug-in vLLM officiel ne mine que sur des lots de préremplissage (prefill) d'au moins 1 024 tokens et jamais sur les étapes de décodage. Une carte qui sert de l'inférence obtient environ 27 % du hashrate d'un mineur dédié sur un trafic de préremplissage pur, environ 5,4 % sur un mélange conversationnel composé à 80 % de décodage, et seulement environ 1 % sur un trafic dominé par le décodage. Sans demande d'inférence payante, un mineur rationnel multiplie des matrices aléatoires ; le protocole vérifie que les matrices engagées ont été multipliées honnêtement, non qu'elles étaient utiles. **Le théorème tient, l'incitation non.** L'utilité de Pearl est un problème commercial hors chaîne : quelqu'un doit payer pour l'inférence.

CHAPITRE 04

04 PEARL · LES RÈGLES
Pearl : risque de règles, numérique et gouvernance

Quatre forks en deux mois

L'autre visage de Pearl est la vitesse à laquelle ses règles changent. De juin à août 2026, le laboratoire a imposé quatre changements de consensus à des hauteurs codées en dur : un hard fork mixture-of-experts (bloc 71 935, vers le 12 juin), un soft fork réservé aux modèles denses (91 630, vers le 23 juillet), un soft fork de pénalité de rang (96 251, vers le 6 août) et un hard fork à graine salée (99 000, vers le 11 août). Après le fork à graine salée, les anciens mineurs ne pouvaient plus produire que des parts invalides ; la pénalité de rang a réduit de 50 % à 87,5 % le hashrate effectif des configurations de rang élevé.

L'étape suivante est PIP-3, encore à l'état de projet : elle fait passer le minage de l'INT7 au FP8 × FP8 → FP32 et exige des profils de dispositif exacts au bit près, **en ne prenant en charge que deux architectures, Hopper et Blackwell**. Le document qualifie la dépendance au matériel de mal nécessaire des protocoles en virgule flottante et avertit que toute tolérance dans la vérification ouvre la porte aux attaques par grinding. Le 12 novembre, les paramètres mainnet de la branche fp8-scheme ne fixaient aucune hauteur d'activation. Selon le projet actuel, la série RTX 40, l'A100 et les cartes AMD ne sont pas admises, alors qu'un modèle de prIstats à faible niveau de confiance estime qu'environ 82 % du hashrate provient de cartes grand public. Imposer le fork risque de provoquer une scission de la chaîne ou une fronde des mineurs.

Un trilemme du format numérique

Pearl fait face à un trilemme : un format numérique peut être utile pour l'IA (FP8 ou FP4), vérifiable au bit près d'un matériel à l'autre, et neutre vis-à-vis du matériel, mais au plus deux des trois. Le schéma INT7 est vérifiable et neutre, mais diverge de la direction prise par le silicium d'IA. Le FP8 avec liste blanche est utile et vérifiable, mais lie le consensus à un seul fournisseur. Le FP8 avec tolérance est utile et neutre, mais ouvre la porte aux attaques par grinding. L'arithmétique en virgule flottante des cœurs tensoriels ne suit généralement aucune norme numérique et n'est pas reproductible d'une génération à l'autre ; sur Hopper, par exemple, l'accumulation FP8 ne conserve que 13 bits fractionnaires et tronque. Du point de vue d'un concepteur de puces, un « profil Hopper » est une arithmétique entièrement déterminée qu'un ASIC pourrait implémenter au bit près, peut-être plus efficacement que le GPU. Mais ce serait une activité en zone grise, contraire à l'intention affichée du protocole, avec un risque de gouvernance très élevé analyse.

Gouvernance

La gouvernance est concentrée. Le laboratoire a rédigé toutes les propositions de protocole ; les forks s'activent à des hauteurs codées en dur sans signalement des mineurs ; il n'existe pas de seconde implémentation du client ; et début octobre, aucun audit tiers du nœud ni des circuits à divulgation nulle de connaissance n'avait été publié. Le quotidien israélien TheMarker a rapporté que le laboratoire avait levé environ 120–130 millions de dollars lors d'un tour discret (environ 55 millions de dollars en capital plus un fonds d'infrastructure GPU d'environ 75 millions de dollars) ; Pearl ne l'a pas confirmé par ses canaux officiels. « Aucune allocation de tokens aux investisseurs » et « une société dotée d'actionnaires en capital-risque » peuvent être vrais tous les deux, mais la tension demeure : les revenus d'IA peuvent aller au capital tandis que la dilution retombe sur les détenteurs de tokens.

Au total, Pearl est l'expérience de « calcul comme consensus » la plus nette à ce jour : un mécanisme inédit, vérifiable on-chain, avec une unité qui correspond terme à terme au matériel d'IA. Mais l'utilité n'est pas au rendez-vous, et son budget de sécurité, son hashrate et l'enthousiasme de ses mineurs dépendent tous d'un token dont la demande principale reste à former et dont l'offre augmentera d'environ 78 % en 12 mois. Trois signaux relevés par Blockworks méritent d'être suivis : la mise à niveau se déroule-t-elle sans heurts, la part des blocs portant du travail d'IA augmente-t-elle sensiblement, et un fournisseur d'inférence répercute-t-il les revenus du minage sur ses clients.

CHAPITRE 05

CHAÎNES À TRAVAIL UTILE

05 Chaînes à travail utile : ni les matrices ni la divulgation nulle de connaissance n'arrêtent les ASIC

Kaspa : une matrice n'est ni utile ni à l'abri des ASIC

Kaspa est le cas d'école. Son kHeavyHash place un produit matrice-vecteur 64 × 64 sur des valeurs de 4 bits entre deux hachages, et le produit est jeté. Environ 16 mois après son lancement équitable en novembre 2021, IceRiver a annoncé les premiers ASIC Kaspa le 14 mars 2023, et fin 2023 le minage sur GPU était terminé. Comparer une RTX 4090 à 3,8 GH/s et 300 W avec le KS7 de Bitmain à 77 J/TH donne un écart d'efficacité d'environ **1000 fois** estimation. Après le festin est venue la compression : le 13 novembre, le réseau tournait à 358,9 PH/s, environ 64 % sous son niveau de 1 EH/s d'août 2024, et le meilleur KS7 ne rapporte que **0,86 dollar par jour** à 0,10 dollar le kWh.

Aleo : la divulgation nulle de connaissance n'arrête pas non plus le matériel spécialisé

Le puzzle des prouveurs d'Aleo est conçu pour imiter le travail de génération d'une preuve SNARK ; il ne prouve pas les transactions des utilisateurs, et le projet accueille ouvertement les ASIC ZK. Environ cinq mois après son mainnet de septembre 2024, des machines dédiées étaient livrées. Au troisième trimestre 2025, les prouveurs ont reçu 31,2 millions d'ALEO (environ 7,8 millions de dollars) alors que le réseau n'a perçu qu'environ 43 000–44 000 dollars de frais : **un multiple de subvention d'environ 175–180**. Le prix est passé d'environ 9 dollars lors des premières cotations à 0,22 dollar à la fin de ce trimestre, et ARC-46 exige désormais une mise de 100 000 à 2,5 millions d'ALEO par solution, ce qui transforme le minage en seuil de staking.

Qubic et Bittensor

Qubic vend une « preuve de travail utile » qui entraînerait son réseau de neurones Aigarth, mais la production la plus visible de son calcul en 2025–2026 a été les blocs d'autres chaînes. Il a d'abord miné du Monero, faisant passer sa part de moins de 2 % à la mi-mai 2025 à plus de 25 % fin juillet et revendiquant 51 % en août, ce qui a provoqué une réorganisation de six blocs (une autre source en compte 18). Le 1er avril 2026, il est passé au minage de Dogecoin avec des ASIC, dont il consacre les recettes à des rachats et à des destructions de tokens. Aucune preuve indépendante n'atteste l'utilité de son entraînement d'IA, et sa valeur de marché est tombée d'environ 280–300 millions de dollars à environ 50 millions de dollars le 13 novembre. Il a aussi mis au jour une externalité : une chaîne qui rémunère du calcul généraliste crée un grand réservoir de hashrate coordonné et louable, qui peut être retourné contre les chaînes adaptées aux CPU.

Bittensor n'est pas une preuve de travail, mais une notation subjective pondérée par la mise : les validateurs notent les mineurs, et les poids sont écrêtés à la médiane pondérée par la mise. C'est le seul réseau de ce groupe à disposer de revenus externes à grande échelle. L'indice SubConnect, qui ne compte que les clients externes vérifiables, affiche environ **70 millions de dollars par an** début novembre 2026 (environ 35 millions de dollars au 1er septembre), dont 78 % à 82 % proviennent du calcul et de l'infrastructure ; mais aux prix de début novembre, ses émissions dépassent 300 millions de dollars par an, si bien que les revenus n'en couvrent que 9 % à 23 %. Les estimations de revenus varient fortement selon la base, d'environ 17 millions de dollars (Artemis) à 172 millions de dollars en rythme annualisé à partir des totaux autodéclarés. En avril, 67 des 128 sous-réseaux brûlaient

plus de 99 % de leurs incitations, et quand l'équipe de premier plan Covenant est partie sur un différend de gouvernance, le TAO a chuté de 24 % en quelques heures.

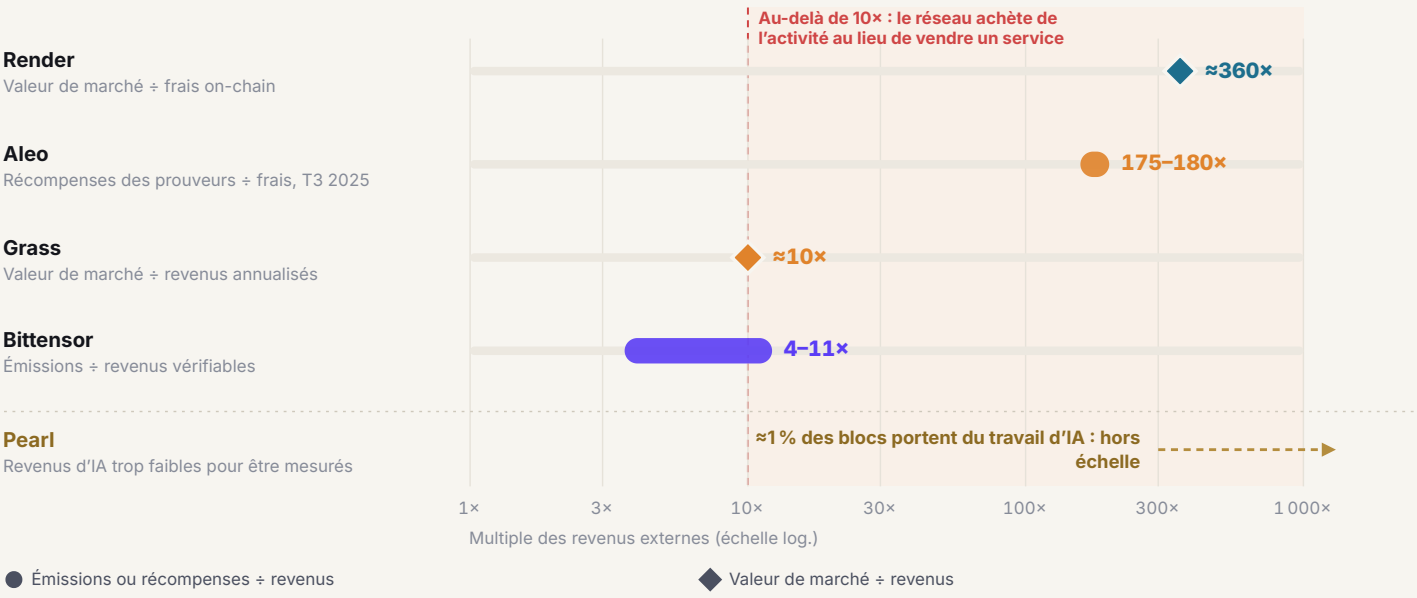
L'histoire donne la même réponse

Les chaînes de nombres premiers de Primecoin étaient rapides à vérifier, mais personne n'en avait besoin. Gridcoin s'appuyait sur des sites de statistiques tiers pour attester les contributions BOINC, et des chercheurs ont forgé des blocs sans effectuer aucun calcul. Les preuves de stockage de Filecoin fonctionnent techniquement, mais au troisième trimestre 2025, le taux d'utilisation n'était que de 36 %, et 99,5 % de ses 793 000 dollars de frais étaient des pénalités, avec des frais de base de seulement 4 100 dollars.

Projet	Travail ou consensus	Valeur de marché (date)	Trajectoire matérielle	Demande payante face à la subvention	Jugement
Pearl	Multiplication matricielle bruitée (MAC INT7 plus certificat ZK)	Environ 500 millions de dollars (13 novembre)	GPU ; le projet FP8 placerait Hopper et Blackwell sur liste blanche	Environ 1 % des blocs avec du travail d'IA ; revenus des mineurs surtout de la subvention	Le mécanisme le plus crédible ; demande non prouvée
Kaspa	kHeavyHash (un produit matrice-vecteur 64 × 64 sur 4 bits entre deux hachages)	Environ 840 millions de dollars (13 novembre)	ASIC environ 16 mois après le lancement ; aujourd'hui tout ASIC	Ne revendique aucune utilité ; revenus des mineurs environ 18 millions de dollars par an estimation	Une matrice n'est ni utile ni à l'abri des ASIC
Aleo	Puzzle SNARK synthétique plus preuve d'enjeu	Environ 118 millions de dollars (fin du T3 2025)	Machines dédiées après environ cinq mois	Récompenses des prouveurs environ 175–180 fois les frais	Travail synthétique, porté par la subvention
Qubic	« uPoW » (prétend entraîner Aigarth)	Environ 50 millions de dollars (13 novembre)	CPU et GPU ; ASIC pour Dogecoin depuis avril 2026	Aucune preuve indépendante d'utilité ; calcul utilisé pour du minage extérieur et des rachats	Agrégation de calcul enveloppée dans un récit d'IA
Bittensor	Consensus Yuma (notation pondérée par la mise)	Environ 2,55 milliards de dollars (11 novembre)	GPU généralistes ; pas de trajectoire ASIC	Environ 70 millions de dollars par an de revenus vérifiables contre plus de 300 millions de dollars d'émissions	Le seul doté de revenus externes à grande échelle ; encore une subvention de 4–11 fois
Filecoin (référence)	Preuves de réplication et d'espace-temps	Environ 1,5 milliard de dollars (fin du T3 2025)	—	Utilisation 36 % ; frais de base trimestriels de 4 100 dollars	Utile, mais en manque de demande

FIG. 04 **Le multiple de subvention : combien de fois ses revenus externes un réseau verse**

Émissions ou valeur de marché divisées par les revenus payés depuis l'extérieur du réseau, échelle log.



Source: Blockworks (Aleo T3 2025), DefiLlama (frais de Render), CoinGecko, hoge.gg (Bittensor), Dealroom (Grass), Blockworks 0xResearch (Pearl) ; calcul de l'auteur.

La leçon pour le matériel : un travail fixe et régulier rémunéré dans un token liquide voit arriver des ASIC en 3 à 16 mois environ, souvent conçus par des challengers avant les acteurs établis. Les prétentions à l'« utilité » n'ont tenu que là où un client extérieur paie pour le résultat. Et les principaux facteurs de matériel rendu inutilisable sont une subvention élevée par rapport aux revenus et le contrôle de la fonction de travail par les développeurs.

CHAPITRE 06

LE PAYSAGE ÉLARGI

06 Le reste de l'IA × crypto : une demande réelle mais faible, et une valeur qui passe aux actions

Marchés de calcul décentralisés

Le calcul décentralisé a des clients payants, mais la demande est faible, concentrée et consiste surtout en revente d'inférence sur des modèles ouverts. Les frais vérifiables on-chain sont modestes : DefiLlama indique environ 2,35 millions de dollars de frais pour Render sur l'année écoulée et environ 680 000 dollars pour Livepeer, et l'ensemble du secteur DePIN n'a gagné qu'environ 72 millions de dollars de revenus on-chain en 2025. Les chiffres autodéclarés sont plus élevés mais définis de façon lâche : Aethir déclare « plus de 127,8 millions de dollars » de revenus pour l'exercice 2025 sans préciser s'il s'agit de facturation brute ou de la part du protocole. Valeur des tokens et revenus ont divergé dans les deux sens : la valeur de marché de Render vaut environ 360 fois ses frais on-chain, tandis que celle d'Aethir, environ 87,7 millions de dollars, est inférieure à ses revenus autodéclarés estimation.

L'argument d'un coût inférieur de 70 % à 85 % ne tient que face aux prix à la demande des hyperscalers. Le 3 novembre, le prix moyen pondéré d'un H100 sur Akash était d'environ **2,73 dollars par heure-GPU** (fourchette 2,04–3,16 dollars), sous les quelque 6,88 dollars d'AWS et 11,06 dollars de Google Cloud, mais au-dessus d'environ 1,73 dollar sur Vast.ai, 1,99 dollar sur le cloud communautaire de RunPod et environ 2,44 dollars pour une réservation d'un an (prix des autres plateformes relevés le 12 septembre ; Akash, l'éditeur, est partie intéressée). La demande est surtout de l'inférence banalisée distribuée via OpenRouter, environ 1,7 milliard de tokens par jour sur AkashML et plus de 4 milliards sur io.net, en concurrence frontale avec tous les néoclouds. Les plus gros montants du « calcul décentralisé » se recentralisent : Axe Compute, qui détient le token d'Aethir comme actif de trésorerie, a signé un contrat de cinq ans d'une valeur de 1,5 milliard de dollars pour louer son propre cluster de 9 200 B300, face à un chiffre d'affaires de 35 000 dollars au premier trimestre 2026.

Entraînement et inférence décentralisés

L'entraînement via Internet a été démontré entre 8 et 40 milliards de paramètres. Pluralis-8B s'est entraîné pendant 40 jours sur 330 nœuds, surtout des GPU grand public, et a atteint 63 % de l'efficacité d'une référence H100 centralisée, même si son article admet que l'exécution n'était pas décentralisée et nécessitait les propres nœuds de Pluralis ainsi qu'une autorité d'admission. Psyche, de Nous Research, a mené à terme le préentraînement Consilience de 40 milliards de paramètres. À l'échelle de la frontière, le travail s'est recentralisé : INTELLECT-3 de Prime Intellect, un modèle mixture-of-experts de 106 milliards de paramètres, a été entraîné pendant deux mois sur 512 H200 reliés par InfiniBand à 400 Gbit/s. Les contraintes sont claires : les nœuds ont besoin d'au moins 200 Mbit/s de bande passante et d'une compression des communications d'environ 100 fois, et la confiance comme la vérification dépendent encore des rôles privilégiés de l'opérateur.

L'argent est allé à des sociétés par actions qui vendent des produits centralisés. Prime Intellect affichait plus de 100 millions de dollars de revenus annualisés avant de lever 130 millions de dollars sur une valorisation de 1 milliard de dollars en juillet ; Venice, une application d'inférence privée, déclare plus de 70 millions de dollars de revenus annualisés et des bénéfices depuis le premier trimestre 2026 (chiffres de l'entreprise), a levé 65 millions de dollars sur une valorisation de 1 milliard de dollars et fonctionne sur des GPU loués. Côté tokens, le réseau d'en-

traînement central de Gensyn est toujours « à déterminer » et son token vaut environ 17,2 millions de dollars ; le token de Ritual, une couche 1 native IA, a chuté de 72 % par rapport à son plus haut après le mainnet (titre seulement).

IA vérifiable

La production a choisi les environnements d'exécution de confiance (TEE). Le calcul confidentiel de Hopper ajoute moins de 7 % de surcoût aux requêtes typiques de LLM et presque rien pour les grands modèles ; le surcoût de production de Phala est d'environ 0,5 % à 5 %, mais fin 2025, 398 de ses 10 004 utilisateurs payaient. La racine de confiance d'un TEE est le fabricant de la puce : TEE.fail, divulgué en octobre 2025, a utilisé un interposeur de bus DDR5 coûtant moins de 1 000 dollars pour extraire les clés d'attestation de machines Intel TDX et AMD SEV-SNP entièrement à jour, ce qui fragilise à son tour le calcul confidentiel sur GPU de NVIDIA, et les deux fabricants déclarent que les attaques physiques sortent de leur modèle de menace. C'est exactement le cadre du DePIN, où les opérateurs de nœuds ont un accès physique. Le zkML peut désormais prouver un LLM complet : DeepProve de Lagrange prouve GPT-2 à environ 174 tokens par minute et Gemma 3 à environ 86, à des ordres de grandeur de l'inférence en production et sans usage payant déclaré.

Paiements d'agents et données

x402 est le meilleur exemple d'avertissement. TRM Labs compte environ 198,9 millions de transactions et 52,7 millions de dollars réglés par les facilitateurs connus depuis mai 2025, à 99,6 % en USDC ; après exclusion des autopaiements, des flux groupés et des vendeurs comptant moins de dix acheteurs, la part réellement initiée par des agents n'est que de 0,6 % à 7,5 %, **environ 5 000–11 000 dollars par mois**. Artemis estime que plus de 95 % de l'activité relève de la signalisation de protocole plutôt que d'achats et de ventes. Pendant ce temps, les rails sont standardisés par les acteurs établis : la x402 Foundation a été créée sous l'égide de la Linux Foundation le 14 juillet, avec Visa, Mastercard, Stripe, Google, AWS, Cloudflare et Coinbase comme membres fondamentaux, et Tempo, qui a coécrit le Machine Payments Protocol avec Stripe, a lancé son mainnet en mars sur une valorisation de 5 milliards de dollars. Les plateformes de tokens d'agents fonctionnent encore selon une logique de launchpad : Virtuals déclare 479 millions de dollars d'« aGDP », et le fondateur d'ElizaOS a déclaré son token mort après le règlement d'une action collective. La valeur a plus de chances d'être captée par les émetteurs de stablecoins et les acteurs établis du paiement ; le rôle distinctif de la crypto pourrait ici être l'identité et la responsabilité des agents.

Les données sont l'un des rares sous-secteurs à disposer de vrais acheteurs hors chaîne. Grass déclare 17 millions de dollars de revenus de données en 2025 et 17 millions de plus au premier semestre 2026, vise 75 millions de dollars pour l'année et se dit rentable ; GRASS vaut environ 328 millions de dollars, soit environ dix fois ses revenus annualisés du premier semestre estimation. Son avantage concurrentiel est l'échelle de collecte que lui donne la bande passante résidentielle distribuée, pas la crypto elle-même. Les DAO de données et les places de marché de données manquent de demande payante visible : Vana vaut environ 21,5 millions de dollars et Ocean environ 27,6 millions de dollars, et Ocean est en litige avec Fetch.ai. La preuve de personnalité gagne de vraies intégrations à mesure que les agents se multiplient : World dit avoir vérifié près de 18 millions de personnes, et Zoom, Docusign et Tinder prennent en charge certaines parties de World ID.

Le capital a voté

Secteur	Demande payante vérifiable	Signaux de subvention et de valorisation	Jugement
Chaînes à travail utile	Bittensor environ 70 millions de dollars par an de revenus externes vérifiables ; Pearl environ 1 % des blocs avec du travail d'IA	Aleo 175–180 fois ; Bittensor 4–11 fois ; Pearl surtout de la subvention	Les mécanismes ont de la valeur ; l'utilité a besoin de clients payants

Secteur	Demande payante vérifiable	Signaux de subvention et de valorisation	Jugement
Marchés de calcul décentralisés	Frais on-chain d'environ 0,68–2,35 millions de dollars par an (Livepeer, Render) ; autodéclarés 4–128 millions de dollars	Valeur de marché de Render environ 360 fois les frais	Réel mais faible, surtout de la revente d'inférence banalisée
Entraînement et inférence décentralisés	Sociétés par actions : Prime Intellect plus de 100 millions de dollars par an, Venice plus de 70 millions (infrastructure centralisée)	Mainnet d'entraînement de Gensyn non lancé ; token de Ritual −72 % depuis son plus haut	Progrès techniques réels ; les profits vont aux actions
IA vérifiable	TEE en production (398 utilisateurs payants chez Phala) ; aucun zkML payant déclaré	—	TEE utiles ; zkML précoce
Paiements d'agents	Flux initiés par des agents environ 5 000–11 000 dollars par mois	Environ 200 millions de transactions, surtout de la signalisation et de la spéculation	Les rails d'abord, la demande plus tard
Réseaux de données	Grass 17 millions de dollars au S1 2026 (autodéclaré)	GRASS environ 10 fois ses revenus annualisés	L'un des rares secteurs avec de vrais acheteurs

Mettez les six secteurs côte à côte et la conclusion est constante : **la subvention est un carburant pour un démarrage à froid, pas un modèle économique.** Le 13 novembre, la catégorie IA de CoinGecko valait environ 18,6 milliards de dollars au total, mais environ 56 % de ce montant correspond à des couches 1 généralistes comme NEAR et ICP, ainsi qu'à des projets au niveau du consensus comme TAO, le TAO staké (peut-être compté deux fois avec TAO) et Pearl ; les tokens d'IA des couches protocole et application totalisent environ 8 milliards de dollars ou moins estimation. C'est à peu près la valorisation cumulée de quatre sociétés par actions : Tempo (5 milliards de dollars), Nous (1,5 milliard de dollars selon la presse), Prime Intellect (1 milliard de dollars) et Venice (1 milliard de dollars). Le capital-risque crypto plafonne à environ 4,5–5 milliards de dollars par trimestre, la hausse venant de l'infrastructure liée à l'IA, dont la part des montants divulgués est passée de 8,1 % à 18,2 % au troisième trimestre 2026 ; et 84,7 % des 118 tokens lancés en 2025 s'échangent sous leur prix de lancement. En une phrase : **la crypto assure la distribution et le règlement, les actions captent les profits**, du moins en 2026.

CHAPITRE 07

07

ÉCONOMIE DES ASIC

La question des ASIC : une formule, trois critères

Une formule : pourquoi l'efficacité cesse de payer au-delà de k = 5

Soit R le revenu quotidien par unité de calcul (revenus des mineurs du réseau divisés par le hashrate du réseau), c le coût quotidien d'une unité de calcul sur le matériel marginal (sur Pearl, un GPU), et k le nombre de fois qu'un ASIC est moins coûteux que ce matériel marginal (c_ASIC = c/k). Tant que les GPU restent les mineurs marginaux, la liberté d'entrée et de sortie pousse R vers c, si bien qu'un ASIC gagne R − c/k par unité de calcul :

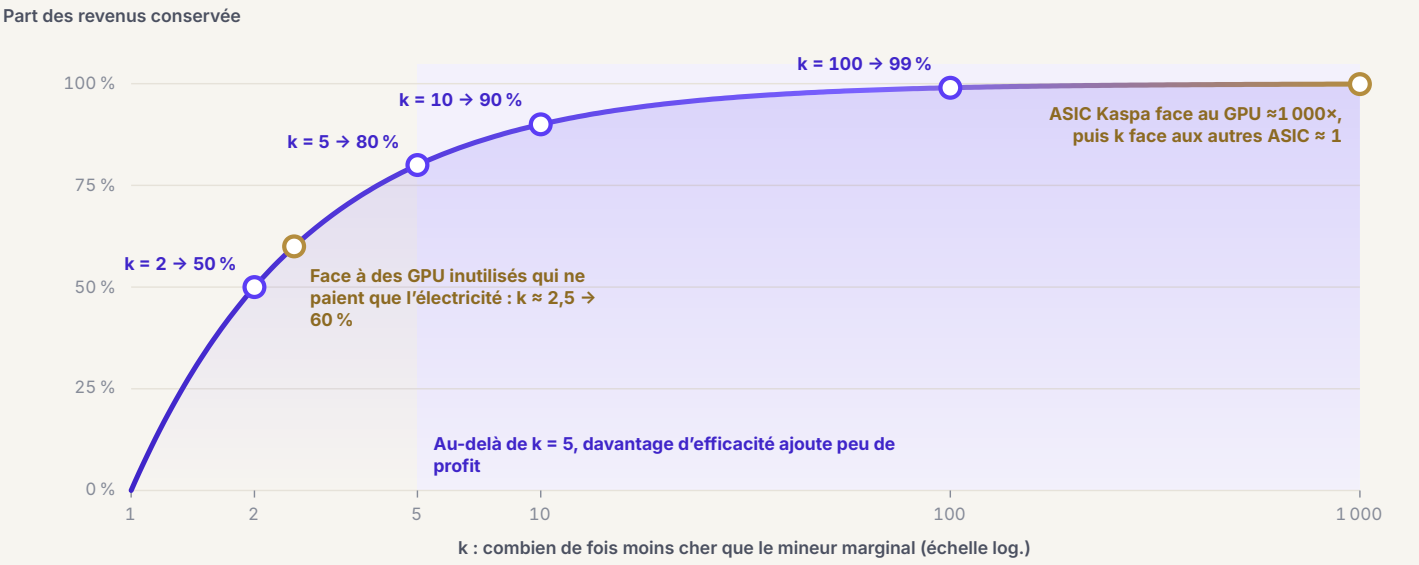
Plafond de profit d'un ASIC ≈ R × (1 − 1/k)

À l'échelle du réseau, le plafond de marge brute annuelle d'une flotte d'ASIC vaut à peu près les revenus des mineurs du réseau × (1 − 1/k), et seulement jusqu'à ce que les ASIC remplacent les GPU comme mineurs marginaux.

La formule a trois conséquences. D'abord, (1 − 1/k) est concave : à k = 2, un ASIC conserve 50 % des revenus, à k = 5 80 %, à k = 10 90 %, à k = 100 99 %. **Au-delà de k = 5–10, davantage d'efficacité n'ajoute presque aucun profit** ; c'est la taille de R et de la cagnotte qui décide. Ensuite, la référence se réinitialise : dès que les ASIC sont les mineurs marginaux, k doit se mesurer face aux autres ASIC. Les ASIC de Kaspa avaient un avantage d'efficacité d'environ 1 000 fois sur les GPU, qui n'a acheté qu'une courte manne ; aujourd'hui, le meilleur KS7 rapporte 0,86 dollar par jour. Enfin, sur une chaîne adaptée aux GPU comme Pearl, la référence n'est souvent pas le prix d'achat d'un GPU mais **l'électricité d'un GPU inutilisé** : une RTX 5090 déjà possédée, à 525 W et 0,10 dollar le kWh, coûte environ 1,26 dollar par jour en électricité et rapporte environ 5–6 dollars estimation. Face à des GPU qui sont des coûts irrécupérables, même un ASIC qui consomme un cinquième de l'énergie n'a qu'un k effectif d'environ 2,5 si son amortissement égale son électricité, et un plafond de profit d'environ 60 % de R (hypothèse illustrative).

FIG. 05 **Pourquoi l'efficacité cesse de payer : un ASIC conserve au plus 1 − 1/k des revenus**

Part des revenus qu'un ASIC peut conserver face au mineur marginal, selon son avantage de coût k



Source: Modèle de l'auteur : quand le matériel marginal fixe le prix, le revenu par unité de calcul tend vers son coût ; chiffres Kaspa et RTX 5090 tirés du rapport.

Trois critères : la cagnotte, l'échelle d'amortissement et les règles

Le premier critère est **la cagnotte**. En annualisant les quelque 2 893 BTC de récompenses de bloc de la première semaine de novembre (environ 222 millions de dollars), les mineurs de bitcoin gagnent environ 11,5 milliards de dollars par an estimation. Les mineurs de Pearl gagnent environ 0,46–0,49 milliard de dollars (aux prix actuels) ; les récompenses des prouveurs d'Aleo, annualisées à partir du troisième trimestre 2025, environ 31 millions de dollars ; et les mineurs de Kaspa, avec une récompense de bloc de 1,945 KAS et dix blocs par seconde, environ 18 millions de dollars estimation. **Une année entière de revenus des mineurs de Kaspa est inférieure au bas de la fourchette du budget de conception d'une seule puce en 5/4 nm (environ 30 millions de dollars).**

Le deuxième critère est **le hashrate nécessaire pour amortir les NRE** :

$$h^* = \mathrm{NRE} \div [R \times (1 - 1/k) \times T \times p]$$

où T est la fenêtre utile en jours et p la probabilité que les règles restent inchangées durant cette fenêtre. Pour Pearl, prenons R = 14 dollars par PH/s et par jour, k = 5, T = 365 jours et des NRE de 20 millions de dollars (une puce à réseau répétitif en N7/N6 ; au bas des fourchettes de 5–30 millions de dollars pour les nœuds intermédiaires selon Silicon Analysts et de 15–100 millions de dollars pour les ASIC de calcul en 7/6 nm selon AnySilicon, car une puce de minage utilise peu d'IP et une structure répétée ; c'est mon hypothèse). Avec p = 1, il faut environ **4,9 EH/s** en fonctionnement pendant une année entière, soit environ 5 % du réseau ; avec p = 0,5, environ **9,8 EH/s**, soit environ 11 % ; et si R retombe au niveau estival d'environ 10 dollars, même p = 1 exige environ 6,9 EH/s estimation. L'échelle n'est pas hors d'atteinte. Le problème, ce sont p et T.

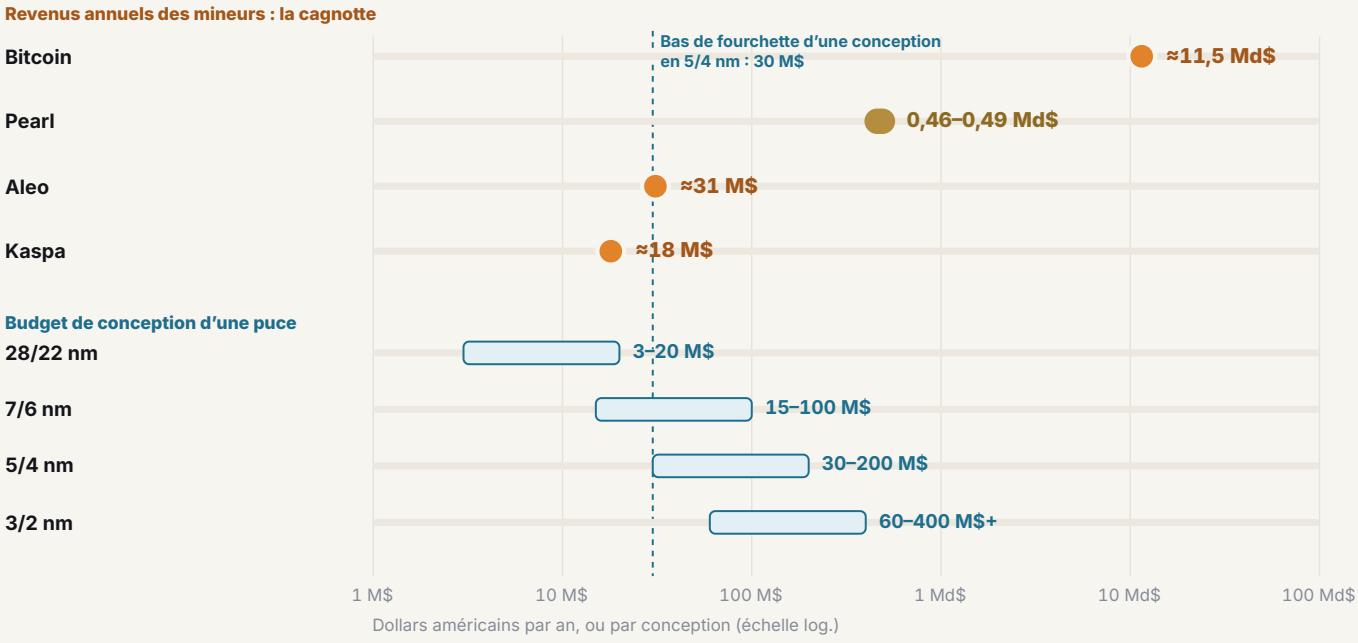
Le troisième critère est **le risque de règles**, et il est souvent décisif. Le fork anti-ASIC de Monero en 2018 a rendu inutilisable le X3 de Bitmain et a brièvement coûté au réseau près de la moitié de son hashrate. Le fork de Sia la même année a mis hors service les mineurs de Bitmain et d'Innosilicon, tandis que le prix d'occasion du mineur Obelisk, lié aux développeurs, est passé de 2 499 dollars à plus de 5 000 dollars. Un peu plus d'un mois après le Merge d'Ethereum, moins de 20 % des anciens mineurs d'Ethereum avaient trouvé un nouveau travail. **Qui contrôle la fonction de travail contrôle le sort des investisseurs en matériel**, et Sia montre que la « résistance aux ASIC » peut devenir l'avantage concurrentiel des développeurs eux-mêmes. Le matériel doté d'un second dé-

bouché survit : après le Merge, les GPU sont passés à l'IA et au cloud, tandis que les ASIC mono-algorithme se sont entassés sur Ethereum Classic. La fenêtre se réduit aussi : le premier ASIC de Kaspa a mis environ 16 mois, celui d'Aleo environ cinq, et le premier à atteindre la frontière d'efficacité est souvent un challenger plutôt qu'un acteur établi.

Réseau	Revenus annuels des mineurs (environ)	Matériel marginal ou de référence	Stabilité des règles	k accessible à un nouvel entrant	Jugement
Bitcoin	11,5 milliards de dollars	Bitmain S23 XP Hyd (8,9 J/TH, 25,5 dollars par TH)	Très élevée	≤1	Une grande cagnotte, mais k appartient aux acteurs établis
Pearl	0,46–0,49 milliard de dollars (aux prix actuels)	GPU possédés ou inutilisés (coût marginal ≈ électricité) et GPU de centre de données	Très faible : quatre forks en deux mois ; liste blanche FP8 en suspens	Peut-être 3–10 fois en efficacité analyse ; moins face aux GPU à coût irrécupérable	La cagnotte est assez grande ; le risque de règles y oppose son veto
Aleo	Environ 31 millions de dollars (T3 2025, annualisé)	IceRiver AE2 et autres machines dédiées	Moyenne : seuil de staking ARC-46	≈1	Une petite cagnotte, et le seuil monte
Kaspa	Environ 18 millions de dollars	KS7 (77 J/TH)	Élevée	≈1	Revenus annuels inférieurs au bas de la fourchette d'un budget de conception en 5/4 nm

FIG. 06 **La plupart des cagnottes sont inférieures au budget de conception d'une seule puce**

Revenus annuels des mineurs de quatre réseaux face au coût de conception d'une puce par nœud, échelle log.



Source: Hashrate Index (bitcoin, novembre 2026), prlstats et Kryptex (Pearl), Blockworks (Aleo, T3 2025), minerstat (Kaspa), AnySilicon (budgets de conception, 2026) ; annualisation de l'auteur.

Le quadrant « grande cagnotte, règles stables et k > 1 » est vide aujourd'hui : bitcoin a la cagnotte et des règles stables, mais son k appartient aux acteurs établis ; la cagnotte de Pearl est assez grande, mais ses règles sont les moins stables ; Kaspa et Aleo ont des règles assez stables et des cagnottes trop petites.

CHAPITRE 08



08

PUCE PAR PUCE

Cinq types de puces, cinq réponses

PoUW à la Pearl : pas de puce INT7 dédiée ; parier sur un mineur capable de servir de l'inférence

Les arguments en faveur d'un ASIC sont réels. Seuls environ 1 % des blocs portent aujourd'hui du travail d'IA, si bien que l'argument économique du livre blanc, selon lequel un ASIC dédié perd la capacité d'effectuer un travail utile et gagne donc moins, est au plus faible en ce moment. Les éléments les plus coûteux d'un GPU, la HBM et le packaging CoWoS, représentent environ 63 % du coût de fabrication d'un H100, le die logique seulement environ 9 %. Une puce qui ne fait que des multiplications-accumulations sur entiers de 7 bits, garde ses opérandes dans la SRAM embarquée et fonctionne à basse tension pourrait plausiblement atteindre 3 à 10 fois l'efficacité énergétique par opération d'un GPU du même nœud, et un avantage plus grand en coût par opération [analyse ; pas encore de mesures]. Les arguments contre sont plus solides. Les délais des fondeurs aux nœuds de pointe peuvent dépasser 50 semaines, et une nouvelle équipe a besoin d'environ 18 à 24 mois du démarrage à la production en volume estimation ; le temps qu'une puce arrive, Pearl pourrait fort bien être passé au FP8, puis au FP4. Et comme PIP-3 définit l'arithmétique admise au moyen d'une liste blanche de dispositifs, une puce INT7 cesserait tout simplement de fonctionner dès son activation.

Ma conclusion est donc : **ne pas lancer maintenant le tape-out d'une puce Pearl INT7 dédiée.** Il existe deux meilleurs paris. Le premier est un accélérateur conçu d'abord pour l'inférence et capable de miner : bâti autour d'un GEMM INT8/FP8 généraliste et d'opérandes résidant en SRAM, de sorte que, si l'accès au minage est refusé, la valeur se reporte sur un vrai marché de l'inférence. Le second est de participer au débat sur le protocole et de plaider pour une arithmétique canonique neutre vis-à-vis des fournisseurs, qu'un ASIC peut implémenter exactement, assortie d'un engagement de stabilité de l'algorithme, condition préalable à l'existence du quadrant « grande cagnotte, règles stables, k > 1 ». D'ici là, une architecture peut être validée sur FPGA ou sur wafers multi-projets (MPW) en nœud mature, en maintenant le risque de NRE de l'ordre du million de dollars. Le chapitre 10 traduit cela en budget par étapes.

Mineurs à la bitcoin : les acteurs établis possèdent k, et l'électricité est l'actif rare

Le risque de règles du bitcoin est le plus faible, mais la concurrence pousse le k d'un nouvel entrant sous 1. Le S23 XP Hyd de Bitmain atteint 8,9 J/TH, se vend 25,5 dollars par TH et est livré à partir de novembre ; l'efficacité des modèles phares est passée de 17,5 J/TH pour le S21 en 2023 à 8,9 J/TH, soit environ 49 % en trois ans. Un nouvel entrant livrant en volume en 2027–2028 aurait besoin d'environ 8 J/TH ou mieux pour tenir estimation. Les challengers s'en vont : Canaan a annoncé pour le deuxième trimestre 2026 un chiffre d'affaires de 31,9 millions de dollars avec une marge brute négative de 29,3 millions de dollars, et n'a prévu que 11–15 millions de dollars pour le troisième trimestre ; les ventes par Bitdeer de ses propres machines à des acheteurs extérieurs sont tombées à 0,3 million de dollars au troisième trimestre, avec une marge brute du groupe de 3,4 % ; Auradine s'est rebaptisé Velaura AI et s'est tourné vers les puces d'IA ; et Core Scientific a annulé une commande d'environ 15

EH/s de machines en 3 nm auprès de l'unité Proto de Block, en comptabilisant une perte de 41,9 millions de dollars.

La demande se contracte aussi. Le 9 novembre, le hashprice était d'environ 31,89 dollars par PH/s et par jour, la courbe à terme jusqu'en avril 2027 était en moyenne d'environ 32,7 dollars, et les frais ne représentaient que 0,72 % des récompenses de bloc. À 0,045 dollar le kWh, le S23 XP Hyd s'amortit en environ 3,1 ans estimation, soit à peu près au moment du prochain halving attendu vers 2028. Plus important encore est le coût d'opportunité de l'électricité : les baux pour l'IA et le calcul haute performance rapportent désormais environ 1,67–1,9 million de dollars par mégawatt contracté et par an, ce qui, avec un résultat opérationnel net de 80 % à 85 %, représente environ 1,5 à 2,9 fois la marge brute du minage par mégawatt (environ 0,56–0,9 million de dollars) [estimation ; les deux bases de mégawatts diffèrent]. Core Scientific cessera de miner au quatrième trimestre ; Bitfarms, devenu Keel, a fermé tous ses sites de minage aux États-Unis ; IREN a gagné 200 millions de dollars grâce au cloud d'IA et seulement 28 millions de dollars grâce au minage en un trimestre. **L'électricité et la capacité des centres de données sont les actifs ; les machines de minage sont des consommables.**

L'arithmétique des NRE est tout aussi implacable. L'efficacité de pointe exige désormais des procédés de la classe 3 nm (selon des sources secondaires, MicroBT utilise le SF3 de Samsung, et les puces d'Auradine et de Proto sont en 3 nm). Les masques coûtent environ 10–20 millions de dollars et les NRE complètes 30–100 millions de dollars ou plus ; à 20–25 dollars par TH et une marge brute de 30 %, chaque tranche de 1 EH/s vendue rapporte environ 6–7,5 millions de dollars de marge brute, si bien que couvrir une R&D annuelle de l'ordre de celle de Bitdeer, environ 110–160 millions de dollars, exige de vendre environ 20–25 EH/s par an à des acheteurs extérieurs, soit environ 2 % du réseau estimation. La conclusion : **une nouvelle équipe ne doit pas entrer sur le marché ouvert des machines de minage bitcoin.** Trois cas gardent un sens : l'intégration verticale avec une électricité bon marché en propre ; des formes de niche comme les mineurs domestiques et de chauffage (Canaan lance aussi des produits de chauffage domestique) ; et la participation aux feuilles de route des acteurs établis par l'IP et les services de conception.

Puces d'entraînement d'IA : pas pour une petite équipe

Pour les puces d'entraînement, k est le plus difficile à obtenir. À génération et précision égales, l'Ironwood de Google ne devance le B300 en efficacité FP8 au niveau de la puce que d'environ 1,1 à 1,8 fois (la puissance d'Ironwood est une estimation de journaliste) estimation ; les 30 à 80 fois souvent cités d'après l'article sur le TPU v1 ont été mesurés face à un K80, dépourvu de cœurs tensoriels. Le vrai levier est le format numérique : passer du FP8 au FP4 donne généralement 2 à 3 fois le débit sur le même silicium, et c'est là que NVIDIA avance le plus vite : il revendique 50 PF d'inférence NVFP4 pour Rubin, en volume au second semestre 2026. Les seuils de coût et d'approvisionnement sont plus élevés encore : les budgets de conception en 3/2 nm vont de 60 à 400 millions de dollars ou plus ; la capacité CoWoS en 2026 est d'environ 106 000–130 000 wafers par mois, et TSMC envisagerait des hausses de prix allant jusqu'à 20 % ; à un prix contractuel de 15,6 dollars par Go pour la HBM3E, la HBM seule d'un accélérateur de 288 Go coûte environ 4 500 dollars estimation. Ce marché appartient au codéveloppement à la Broadcom avec les hyperscalers (chapitre 11). La conclusion : **les petites équipes ne doivent pas concevoir de puces d'entraînement autonomes** ; qu'elles fassent plutôt de l'IP, des chiplets ou des services de conception.

Puces d'inférence d'IA : échanger la généralité contre k, et c'est payant

L'inférence est le domaine où les ASIC ont le plus de marge, car k peut s'acheter en renonçant à la généralité. Etched affirme que son serveur à huit puces fait tourner Llama-70B à plus de 500 000 tokens par seconde, environ 20 fois huit H100 ; Taalas grave en dur les poids du modèle dans une ROM de masque et revendique une consommation 10 fois plus faible et un coût de fabrication 20 fois plus faible (dans les deux cas selon le fournisseur). Les architectures centrées sur la SRAM évitent la HBM et le CoWoS, et leurs nœuds sont plus abordables : Taalas utilise

le 6 nm de TSMC, et Groq est passé du 14 nm au 4 nm de Samsung. La distribution crypto n'aide pas les puces d'inférence, puisque les réseaux de calcul décentralisés tournent presque entièrement sur NVIDIA ; les portes d'entrée sont l'économie des achats des néoclouds et le vaste marché intérieur chinois de l'inférence, où les centres de données financés par l'État doivent utiliser des puces d'IA nationales, où les importations de H200 sont freinées et où Pékin veut que les grands groupes Internet fassent tourner l'inférence sur des puces nationales. La conclusion : **cela vaut la peine d'en concevoir**, à condition d'avoir un client clair et que le degré de verrouillage aux modèles et à l'architecture soit un choix délibéré. C'est aussi le meilleur second débouché pour une puce de minage PoUW.

Accélérateurs ZK et FHE : se couvrir sur le ZK, se positionner tôt sur le FHE

Les GPU ont déjà rattrapé la preuve ZK en temps réel : SP1 Hypercube de Succinct a prouvé 99,7 % des blocs Ethereum en moins de 12 secondes sur 16 RTX 5090. Cysic affirme que sa puce ZK C1 est 10 à 100 fois plus efficace que les prouveurs sur GPU et sera livrée en 2026, mais sa demande provient largement de ses propres récompenses en tokens selon le fournisseur ; je n'ai trouvé aucune preuve publique que les ASIC du puzzle d'Aleo servent à de vraies preuves ZK. La preuve ZK est entière et déterministe, de forme proche de ce que font les puces de minage, mais les systèmes de preuve changent vite (SP1, RISC Zero, Halo2, Plonky2 et d'autres). La conclusion : **attendre et se couvrir**, avec des moteurs MSM, NTT et de hachage programmables, d'abord sur FPGA, puis en ASIC.

Le FHE est la primitive la plus liée au matériel, avec le plus fort potentiel. Heracles d'Intel, en Intel 3 avec 197 mm², 176 W et 48 Go de HBM3, exécute les opérations FHE 1 074 à 5 547 fois plus vite qu'un Xeon à 24 cœurs, mais Intel n'a pas de projet commercial et le responsable du projet est parti en décembre 2025. Niobium conçoit sur le 8LPU de Samsung avec SemiFive comme partenaire de conception (un contrat de développement d'environ 6,86 millions de dollars) et a levé plus de 23 millions de dollars ; Optalysys, avec une approche photonique, a levé 31 millions de dollars et estime qu'une puce commerciale complète est encore à deux ou trois ans ; et la conception FPGA open source de Zama atteint environ 13 000 bootstraps programmables par seconde à environ 200 W, la référence publique que tout ASIC doit nettement battre. La conclusion : **se positionner tôt et valider avec peu d'actifs**, pour une fenêtre autour de 2027–2029. La demande a plus de chances de s'ancrer dans l'IA chiffrée pour les entreprises et les États que dans les cryptomonnaies elles-mêmes, et le FHE a besoin de HBM, si bien que sa structure de coûts ressemble à celle d'une puce d'entraînement.

Type de puce	D'où vient k	Demande et cagnotte	NRE et seuils	Risque de règles ou de verrouillage	Second débouché	Jugement
Puce de minage PoUW à la Pearl	Suppression de la HBM, du CoWoS et des unités généralistes : peut-être 3–10 fois en J/op analyse	Revenus des mineurs environ 0,46–0,49 milliard de dollars par an, volatils	Réseau répétitif N7/N6 environ 10–30 millions de dollars analyse	Très élevé : le protocole décide du format numérique et de la liste blanche des dispositifs	Inférence d'IA, si conçue en GEMM INT8/FP8 généraliste	Pas de puce INT7 dédiée ; concevoir d'abord pour l'inférence et capable de miner, ou attendre une arithmétique canonique
Puce de minage bitcoin SHA-256	k ≤ 1 face à Bitmain	Environ 11,5 milliards de dollars par an, marché ouvert tenu par deux entreprises	Masques classe 3 nm 10–20 millions de dollars ; NRE complètes 30–100 millions de dollars ou plus	Faible (algorithme stable) ; concurrence et halving : risque élevé	Presque aucun (algorithme unique)	Les nouveaux entrants restent hors du marché ouvert ; seulement électricité propre, formes de niche ou IP

Type de puce	D'où vient k	Demande et cagnotte	NRE et seuils	Risque de règles ou de verrouillage	Second débouché	Jugement
Puce d'entraînement d'IA	Réseaux systoliques de même génération seulement 1,1–1,8 fois ; le format est le levier	Le plus grand (Broadcom environ 58 milliards de dollars de revenus de semi-conducteurs d'IA sur l'exercice 2026)	Conception en 3/2 nm 60–400 millions de dollars ou plus ; CoWoS et HBM	Cadence annuelle de NVIDIA	—	Les petites équipes font de l'IP, des chiplets et des services de conception, pas des puces autonomes
Puce d'inférence d'IA	Renoncer à la généralité : 10–20 fois revendiqué selon le fournisseur	Grand et fragmenté ; le marché intérieur de la Chine continentale est vaste en soi	Architectures SRAM en N6/N7–N4 abordables	Verrouillage aux modèles et à l'architecture	C'est le marché principal	Vaut la peine : signer le client, puis choisir le verrouillage
Accélérateur ZK	10–100 fois les GPU revendiqué selon le fournisseur	Faible, porté surtout par les récompenses en tokens	N7/N6 puis N5 ; FPGA d'abord	Les systèmes de preuve changent vite	Limité hors du ZK	Attendre et se couvrir : moteurs MSM et NTT programmables
Accélérateur FHE	1 074–5 547 fois un Xeon (Heracles)	Pas encore formée ; ancrée dans les entreprises et les États	Exige de la HBM ; structure de coûts proche des puces d'entraînement	Normes et écosystème non stabilisés	Calcul chiffré, IA privée	Se positionner tôt, valider avec peu d'actifs ; fenêtre 2027–2029

FIG. 07

Cinq types de puces, cinq réponses

D'où vient l'avantage de coût k, sur quoi le silicium peut se vendre si les règles changent, et le verdict



Puce de minage PoUW à la Pearl

PAS MAINTENANT

k : sans HBM ni unités généralistes : 3–10× par op. (déduit)

Second débouché : inférence d'IA, si GEMM INT8/FP8 généraliste

Parier sur une puce d'inférence capable de miner, pas sur un composant INT7 dédié



Puce de minage bitcoin SHA-256

S'ABSTENIR

k : ≤1 face aux acteurs établis

Second débouché : presque aucun

Seulement avec électricité propre, forme de niche, ou en IP



Puce d'entraînement d'IA

DE L'IP, PAS DES PUCES

k : 1,1–1,8× à précision égale

Second débouché : —

Packaging et HBM décident ; faire IP, chiplets, services



Puce d'inférence d'IA

CONCEVOIR

k : renoncer à la généralité : 10–20× revendiqué

Second débouché : c'est le marché principal

Signer le client, puis choisir le verrouillage



Accélérateur de preuve ZK

ATTENDRE ET SE COUVRIR

k : 10–100× les GPU revendiqué

Second débouché : limité hors du ZK

Moteurs MSM et NTT programmables ; FPGA d'abord



Accélérateur FHE

SE POSITIONNER TÔT

k : 1 074–5 547× un Xeon (Heracles)

Second débouché : calcul chiffré, IA privée

Une fenêtre 2027–2029 ; valider avec peu d'actifs

Source: Synthèse de l'auteur à partir des chapitres 7 et 8 ; chiffres des fournisseurs signalés comme tels dans le rapport.

CHAPITRE 09

NŒUDS DE GRAVURE

09 Nœuds de gravure : laisser le coût le plus lourd sur la durée de vie choisir

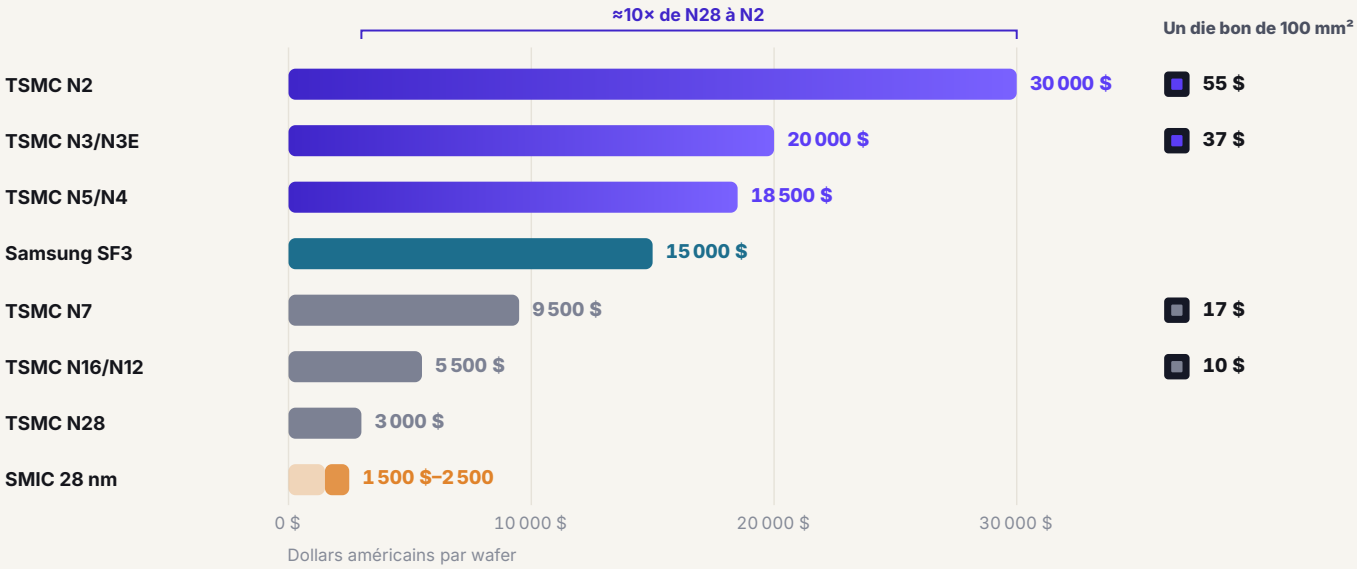
L'échelle des coûts : un wafer de pointe coûte dix fois un wafer mature

Commençons par le grand livre. Un wafer de pointe coûte désormais environ dix fois un wafer mature : TSMC N2 environ 30 000 dollars, N3 environ 20 000 dollars, N7 environ 9 500 dollars et N28 environ 3 000 dollars ; Samsung SF3 environ 15 000 dollars, soit environ 25 % de moins que N3 ; et le coût de fabrication de la classe 7 nm de SMIC serait supérieur de 40 % à 50 % à celui de TSMC. TSMC augmenterait encore ses prix de 3 % à 6 % à partir de janvier 2027, surtout pour le 2 nm et le 3 nm. Les coûts de masques et de conception diffèrent encore davantage, et les sources divergent, si bien que le tableau montre les deux.

Nœud	Prix du wafer (estimation)	Jeu de masques (valeur de modèle / fourchette du guide)	Budget de conception (AnySilicon)	MPW (par projet)
TSMC N2	Environ 30 000 dollars	—	3/2 nm : 60–400 millions de dollars ou plus	—
TSMC N3/N3E	Environ 20 000 dollars (18 000–27 000 dollars)	15 millions de dollars / 10–20 millions de dollars	Idem	—
Samsung SF3	Environ 15 000 dollars (13 000–17 000 dollars)	12 millions de dollars	—	—
TSMC N5/N4	Environ 18 500 dollars (16 000–21 300 dollars, faible confiance)	6,5 millions de dollars / 10–20 millions de dollars	5/4 nm : 30–200 millions de dollars	—
TSMC N7	Environ 9 500 dollars (8 100–10 900 dollars)	3,5 millions de dollars / 5–10 millions de dollars	7/6 nm : 15–100 millions de dollars	Commercial 100 000–500 000 dollars
TSMC N16/N12	Environ 5 500 dollars	1,8 million de dollars	16/12 nm : 6–40 millions de dollars	Académique environ 35 000 dollars par mm²
TSMC N28	Environ 3 000 dollars	1 million de dollars / 1–3 millions de dollars	28/22 nm : 3–20 millions de dollars	Commercial 50 000–150 000 dollars ; académique environ 12 000 dollars par mm²
SMIC 28 nm	1 500–2 500 dollars	—	—	—
SMIC classe 7 nm (N+2/N+3)	40–50 % au-dessus de TSMC	—	—	—

FIG. 08 **Un wafer de pointe coûte dix fois un wafer mature**

Prix estimé d'un wafer de 300 mm par nœud, et coût d'un die bon de 100 mm² à 85 % de rendement



Source: Prix des wafers de Silicon Analysts (estimations, N5 faible confiance), Wccftech (SMIC) ; coûts des dies calculés par l'auteur avec un rendement supposé.

Ensuite, la physique. Les mesures classiques de Horowitz montrent qu'une instruction programmable coûte environ 70 pJ alors que l'opération arithmétique elle-même ne coûte que quelques picojoules : **l'arithmétique est bon marché, le surcoût est cher**, ce qui explique l'existence des réseaux systoliques et des puces dédiées. Diviser par deux la finesse de gravure divise à peu près par deux l'énergie par opération, mais la tension d'alimentation stagne autour de 0,9–1,1 V. Un wafer N2 coûte 50 % à 67 % de plus qu'un wafer N3, si bien qu'à moins d'une hausse de densité supérieure à environ 1,5 fois, le coût par transistor augmente au lieu de baisser estimation : les nœuds de pointe doivent se rentabiliser par l'efficacité et la performance. Pour un die de 100 mm² avec un rendement de 85 %, le coût de chaque die bon est d'environ 55 dollars en N2, 37 dollars en N3, 29–34 dollars en N5, 17 dollars en N7 et 10 dollars en N16 [estimation ; le rendement est une hypothèse].

Choisir un nœud selon le type de puce et l'équipe

Ma règle : **laisser le poste le plus lourd du coût de la puce sur sa durée de vie choisir le nœud.** Pour les puces de calcul qui tournent à pleine charge en continu (minage, ZK), l'électricité domine le coût sur la durée de vie ; une fois l'algorithme stable, elles doivent donc passer en N5/N4 et en N3/SF3 ; tant que l'algorithme n'est pas fixé ou que l'argent manque, il faut d'abord valider sur MPW en nœud mature. Pour les puces d'entraînement, le goulet d'étranglement est le packaging et la mémoire (la HBM et le CoWoS représentent environ 63 % de la nomenclature d'un H100), de sorte que le nœud est secondaire. Pour les puces d'inférence, la première décision est l'architecture mémoire : les conceptions centrées sur la SRAM peuvent rester en N6/N7 à N4, tandis que les conceptions à HBM héritent de la structure de coûts des puces d'entraînement. Le FHE est limité par la bande passante mémoire et ne peut éviter la HBM.

Type de puce	Start-up ou petite équipe	Entreprise à l'échelle	Voie de la Chine continentale	Points clés
Puce de minage PoUW à multiplication matricielle (résidente en SRAM, sans HBM)	Prototype sur MPW 28 nm ou 16/12 nm ; premier volume en N7/N6 ; N5/N4 une fois l'algorithme et les revenus stables ; dies de 50–150 mm² en boîtiers flip-chip standard	N3/N3E ou Samsung SF3	SMIC N+1/N+2 est praticable (ni HBM ni CoWoS nécessaires), pour environ 40–50 % de plus par die bon	L'énergie domine ; les règles du protocole sont le plus grand risque
Puce de minage SHA-256	Ne pas courir après la frontière ; les formes de niche peuvent utiliser des nœuds matures	Classe 3 nm ; les acteurs établis passent du 5 nm au 2 nm de TSMC et du SF3 au 2 nm de Samsung (sources secondaires)	Difficile d'égaler l'efficacité des nœuds de pointe	Le prix de l'électricité et le cycle des halvings déterminent les marges
Puce d'entraînement d'IA	Pas de puce autonome : IP, chiplets, services de conception	N3P puis N2/A16 ; 2–8 chiplets ; sécuriser tôt CoWoS et HBM	SMIC N+2/N+3 avec HBM nationale (le modèle Huawei) ; FP8 par puce environ 0,2–0,5 fois celui de NVIDIA, compensé par l'échelle du système	Le packaging et la HBM sont le goulet d'étranglement
Puce d'inférence d'IA	Centrée sur la SRAM ou câblée en dur, N6/N7 ou N4/N5/SF4 ; INT8, FP8 et FP4 uniquement	N3/N2 avec HBM (long contexte, mixture-of-experts), ou SRAM multi-chiplets	SMIC N+2 avec LPDDR ou GDDR ; la mémoire de classe HBM est le point d'étranglement	Le degré de verrouillage architectural est la décision centrale
Accélérateur ZK	Commencer sur FPGA ou MPW 16/12 nm ; puis N7/N6 et N5	N5/N3	SMIC N+1/N+2 est praticable	Entier et déterministe ; facile à vérifier
Accélérateur FHE	FPGA (face à la référence open source) ou MPW 16/12 nm ; le volume exige de la HBM	N5/N3 avec HBM	Limité par l'accès à la HBM	C'est la bande passante mémoire, et non l'arithmétique, qui fixe la vitesse

Avancé ou mature, quatre types de fondeurs, et chiplets

Option	Principaux atouts	Principales faiblesses	Idéal pour
Nœuds de pointe (N3/N2, SF3/SF2, Intel 18A)	Chaque nœud complet divise à peu près par deux l'énergie par opération ; SRAM dense	Wafers 3–10 fois, masques 5–15 fois et budgets de conception 10–20 fois ceux des nœuds matures ; délais de plus de 50 semaines ; le coût par transistor en N2 pourrait augmenter	Puces de calcul à algorithme stable tournant en continu ; puces d'entraînement
Nœuds matures et intermédiaires (N7/N6, N16/12, N28)	Masques de l'ordre du million de dollars, MPW bon marché, itération rapide, nombreux fondeurs (TSMC, UMC, GlobalFoundries, SMIC)	Efficacité non compétitive à pleine charge en continu	Prototypes, marchés de niche, inférence en périphérie
TSMC	Le meilleur rendement et le meilleur écosystème	Les prix les plus élevés ; capacité N2/N3 tendue ; longues files d'attente pour le CoWoS	Équipes visant la frontière d'efficacité avec des engagements de volume

Option	Principaux atouts	Principales faiblesses	Idéal pour
Samsung Foundry	Wafers environ 20–25 % moins chers (SF3 environ 15 000 dollars contre environ 20 000 dollars pour N3) ; utilisé notamment par la deuxième génération de Groq	L'activité de fonderie perd de l'argent et vise l'équilibre d'ici fin 2027	Puces de calcul sensibles au coût
Intel Foundry	Puces complexes comme Heracles éprouvées en Intel 3	Aucun prix externe public ni donnée client pour 18A ; l'activité de fonderie perd de l'argent	Évaluer l'écosystème et l'accompagnement au cas par cas
Fondeurs de Chine continentale (SMIC et autres)	Proximité d'un vaste marché intérieur ; wafers 28 nm à 1 500–2 500 dollars	Capacité de classe 7 nm d'environ 45 000–60 000 wafers par mois, largement absorbée par Huawei ; rendements d'environ un tiers de ceux de TSMC ; coûts supérieurs de 40–50 % ; DUV uniquement	Puces d'inférence, de minage et ZK pour le marché intérieur
Chiplets	Surmontent les limites du réticule et du rendement ; standard pour les systèmes à HBM (MI355X, Ironwood, Trainium3)	Coût de packaging plus élevé et dépendance au CoWoS ; le test des dies réputés bons (known-good die) ajoute du coût	Puces d'entraînement et grandes puces d'inférence ; inutiles pour les petites puces de minage ou ZK

CHAPITRE 10

BUDGET PAR ÉTAPES

10 Du modèle au silicium : un budget par étapes pour une première puce

Acheter des preuves avant les masques

Une puce full-custom est une série de paris, chacun plus gros que le précédent, et la discipline consiste à acheter des preuves avant d'acheter des masques. J'utilise quatre étapes analyse. L'étape 0 est un modèle : une étude d'architecture qui estime J/op et $/op à partir de l'arithmétique, de la SRAM et des E/S, face au matériel de référence du chapitre 7 ; elle coûte quelques semaines et quelques dizaines de milliers de dollars. L'étape 1 est un prototype FPGA qui exécute la vraie charge de bout en bout, parts de minage ou tokens d'inférence, et valide le chemin de données et l'intégration au protocole ; elle coûte quelques mois et quelques centaines de milliers de dollars. L'étape 2 est un wafer multiprojet : une place commerciale coûte environ 50 000–150 000 dollars en N28 et 100 000–500 000 dollars en N7, et le silicium mesuré remplace les estimations énergétiques du modèle. L'étape 3 est le jeu de masques complet et le premier lot de production : des masques d'environ 1 million de dollars en N28, 3,5 millions en N7 et 15 millions en N3, dans un budget de conception de 15–100 millions de dollars en 7/6 nm. Chaque étape a besoin d'un chiffre qui arrête le projet, comme un k inférieur à 3 face au matériel marginal, ou un hashrate d'amortissement h* que la fenêtre ne permet pas d'atteindre.

FIG. 09 **Acheter des preuves avant les masques : un budget par étapes pour une première puce**

Dépenses cumulées en dollars américains d'un projet N7/N6 sur quatre étapes, échelle log. ; chaque étape a un chiffre qui arrête le projet



Source: Prix des wafers, des MPW et des masques comme au chapitre 9 (Silicon Analysts, AnySilicon, valeurs de modèles) ; calendrier et dépenses des étapes : hypothèses illustratives de l'auteur.

L'escalier est raide. Dans un projet N7/N6, les trois premières étapes coûtent ensemble environ 1,5 million de dollars et environ onze mois ; la quatrième coûte plus de dix fois plus estimation. C'est pourquoi l'ordre compte davantage que le nœud : si le protocole change son arithmétique, ou si l'acheteur se retire, une équipe qui s'est arrêtée à l'étape 2 perd le prix d'un prototype, et non celui d'un tape-out.

Ce que coûte chaque puce, et ce qui doit être vrai

Projet	Étape 2 (prototype)	Étape 3 (nœud et budget de conception)	Un die bon de 100 mm²	Ce qui doit être vrai pour amortir
Accélérateur conçu d'abord pour l'inférence et capable de miner	MPW 16/12 nm ou N7, 0,1–0,5 million de dollars	N6/N7, environ 15–30 millions de dollars estimation	Environ 17 dollars en N7, hors packaging et mémoire	Un client d'inférence signé porte l'essentiel des NRE ; les revenus du minage sont un bonus, pas le scénario de base
Mineur Pearl INT7 dédié (déconseillé aujourd'hui)	MPW 28 nm, 0,05–0,15 million de dollars	N7/N6, environ 20 millions de dollars	Environ 17 dollars en N7	Environ 4,9 EH/s pendant une année entière avec p = 1, ou 9,8 EH/s avec p = 0,5 (chapitre 7)
Accélérateur ZK	FPGA, puis MPW 16/12 nm	N7/N6, puis N5	Environ 10 dollars en N16, 17 dollars en N7	Les systèmes de preuve se stabilisent sur des noyaux MSM et NTT qu'une conception figée peut servir
Accélérateur FHE	FPGA face à la référence open source	N5/N3 avec HBM, 30 millions de dollars ou plus	Dominé par la HBM et le packaging	Un acheteur d'IA chiffrée en volume, qui n'existe pas encore

Deux points commerciaux en découlent. D'abord, l'efficacité la moins chère est architecturale : garder les opérandes en SRAM et se passer de HBM et de CoWoS supprime l'essentiel du coût d'un GPU (environ 63 % de la nomenclature d'un H100) avant même le choix d'un nœud. Ensuite, le temps est un coût : une nouvelle équipe a besoin d'environ 18 à 24 mois du démarrage à la production en volume et les délais aux nœuds de pointe dépassent 50 semaines, si bien qu'un protocole qui change son arithmétique tous les quelques mois peut rendre une puce obsolète avant que l'étape 3 ne soit rentabilisée.

Trois règles pour une petite équipe

Prises ensemble, ces données donnent trois règles analyse. Premièrement, **rester à l'étape 1 ou 2 tant qu'il n'y a ni acheteur nommé ni version de protocole figée** ; ni le prix du token ni le hashrate ne remplacent l'un ou l'autre. Deuxièmement, **acheter k par l'architecture d'abord et par le nœud ensuite** : un réseau résident en SRAM en N7 peut battre un composant généraliste en N3 sur une charge donnée, pour une fraction du coût des masques. Troisièmement, **écrire le hashrate ou le volume d'amortissement, avec p, avant le tape-out**, et fixer le prix des paiements par jalons sur une version figée du protocole, afin que ce soit le client, et non le concepteur, qui porte le risque de règles.

CHAPITRE 11

11 OPPORTUNITÉS
Où sont les opportunités, et ce que je ferais

Puces sur mesure : la demande vient massivement de l'IA

En 2026, la demande de puces sur mesure vient massivement de l'IA. Les revenus de semi-conducteurs d'IA de Broadcom au troisième trimestre de son exercice 2026 ont atteint 16,7 milliards de dollars, en hausse de 221 % sur un an, soit environ 58 milliards de dollars sur l'année, dont 73 % provenant de XPU pour six clients (d'après une transcription automatique de la conférence sur les résultats). Le chiffre d'affaires de Global Unichip au deuxième trimestre a progressé de 128 % sur un an, avec des revenus de production clés en main en hausse de 188 % et une marge brute de 21,5 %. La répartition par nœud du chiffre d'affaires d'Alchip au premier trimestre était de 21 % en 3/2 nm, 63 % en 7/5 nm et 16 % en 12 nm et au-dessus. Les nouvelles commandes de VeriSilicon pour 2026 ont atteint 8,24 milliards de RMB (environ 1,16 milliard de dollars) fin avril, dont 91,37 % pour le calcul d'IA. Le marché compte trois étages : au sommet, des codéveloppements pluriannuels valant des milliards ; au milieu, des projets du 7 nm au 2 nm avec des NRE allant de quelques millions à plus de cent millions de dollars (5–30 millions de dollars aux nœuds intermédiaires, 30–100 millions de dollars ou plus à la pointe) ; et une longue traîne d'ASIC en nœud mature avec des NRE de 0,5–5 millions de dollars. Les clients crypto et minage forment une niche petite et cyclique, mais ce sont aujourd'hui les « premiers clients » les plus rapides et les moins soumis à autorisation qui paient pour du calcul.

Les compétences d'une équipe de puces de minage qui se transfèrent le mieux, le fonctionnement à basse tension proche du seuil, les réseaux de calcul massivement répliqués, l'alimentation et le refroidissement de classe kilowatt, et les tape-outs rapides aux nœuds de pointe, sont exactement ce dont ont besoin les puces d'inférence et les accélérateurs ZK et FHE. Les marchés de capitaux le valorisent : Velaura AI (anciennement Auradine) a levé 110 millions de dollars en août sur une valorisation supérieure à 1 milliard de dollars. Le contre-exemple est l'unité de puces d'IA embarquée de Canaan, qui a réalisé 0,9 million de dollars de chiffre d'affaires sur l'exercice 2024

tout en absorbant 15 % des charges d'exploitation, et qui a été fermée en juin 2025. La différence tient à un accès au marché indépendant, non à la réutilisation des canaux du minage.

Développement d'IA : les heures sont une marchandise ; la valeur est aux jointures

Dans le bas de gamme, les services de développement d'IA sont banalisés : les tarifs typiques sur Clutch sont de 25–49 dollars de l'heure, et de 50–99 dollars aux États-Unis, au Canada et sur des marchés comparables ; une équipe de développement blockchain coûte d'environ 15 000 dollars à plus de 90 000 dollars par mois. Vendre des heures ne procure aucun avantage durable. La valeur est aux jointures entre les domaines : noyaux de calcul et compilateurs pour puces sur mesure, minage intégré au service d'inférence (le plug-in vLLM de Pearl en est un exemple), inférence vérifiable combinant TEE et vérifications ponctuelles ZK, et paiements et identité pour les agents.

Systèmes IA × crypto : signer l'acheteur de calcul avant le token

Pour quiconque conçoit une chaîne ou un réseau de calcul natif IA, la boîte à outils de vérification de 2026 a pris forme. Par ordre de coût par unité de travail, du plus faible au plus élevé : l'échantillonnage statistique (la preuve d'échantillonnage de Hyperbolic), la délégation arbitrée (Verde de Gensyn, qui n'a besoin que d'un seul vérificateur honnête mais exige une exécution exacte au bit près), les TEE (vitesse quasi native, mais confiance dans le fabricant de la puce), la redondance complète et le zkML généraliste. La conception pratique est en couches : TEE ou échantillonnage pour la faible latence, jeux de contestation ou vérifications ponctuelles ZK pour la finalité ; un article de décembre 2025 a simulé le calcul confidentiel du H100 avec des preuves de fraude et des vérifications ponctuelles ZK aléatoires à 99 % du débit centralisé, pour un coût marginal de 0,07 dollar par requête. La philosophie matérielle détermine la sécurité et la distribution : une PoUW native GPU est d'abord protégée par des GPU de cloud louables, si bien que le coût d'une attaque est lié aux prix de location, et Qubic montre qu'une preuve de travail sur matériel généraliste peut être captée en bloc par des incitations extérieures.

Le coût et le temps ne sont pas l'obstacle. Pearl est passé de l'article (avril 2025) au mainnet en environ 12 mois, en forkant btcd, avec un nœud complet, un portefeuille, un client léger et un logiciel de minage dès le lancement. La plupart des audits DeFi coûtent 25 000–100 000 dollars, les audits d'entreprise ou multichaînes 150 000–250 000 dollars ou plus, les circuits ZK de 80 % à 120 % de plus, et les projets exposant une vraie valeur prévoient 150 000–500 000 dollars par an pour la sécurité. Les vrais obstacles sont la distribution et la liquidité. Des fondateurs rapportent que Binance a demandé de 0 % à 3,5 %, et prétendument 8 %, de l'offre de tokens pour une cotation (Binance nie tirer profit des cotations) ; les six projets examinés par Cointelegraph qui avaient recouru à des accords de tenue de marché « prêt de tokens plus option d'achat » ont tous chuté, et le teneur de marché de Movement a vendu environ 38 millions de dollars de tokens le jour de la cotation. Les alternatives sont la liquidité on-chain, comme pour Hyperliquid, qui a dépassé 10 milliards de dollars de valeur de marché sans cotation sur Binance et dont les marchés perpétuels HIP-3 ont dépassé 3,2 milliards de dollars d'intérêt ouvert en juin, et les ventes publiques via de grandes plateformes d'échange, comme lorsque Monad a levé 269 millions de dollars via Coinbase. Une valeur entièrement diluée élevée avec un flottant faible a été le meilleur indicateur d'un déclin après le lancement en 2025–2026.

La conception du token fait partie de la conception du système. L'émission doit suivre la demande payante plutôt que la précéder : dans ce rapport, les réseaux aux multiples de subvention les plus sains lient les récompenses à un travail payé par un acheteur extérieur, et ceux aux pires multiples paient du hashrate ou de la mise sans rien derrière. Une règle pratique consiste à publier, avant le lancement, le multiple de subvention visé par le réseau pour ses première, deuxième et troisième années, et à réduire l'émission lorsqu'il n'est pas atteint analyse.

Direction	Qui achète, et disposition à payer	Fenêtre	Compétences clés	Risque principal
Accélérateur conçu d'abord pour l'inférence et capable de miner	Néoclouds, déploiements privés, réseaux PoUW	2027–2028 (selon les forks FP8 et FP4)	GEMM INT8, FP8 et FP4 ; architecture SRAM ; intégration au protocole	Admission par le protocole, verrouillage architectural
Puce d'inférence SRAM ou câblée en dur	Néoclouds, centres de données en Chine continentale, périphérie	Maintenant	Compilateurs et adaptation des modèles ; conception N6–N4	Les modèles changent trop vite
Puce de minage (niche ou intégration verticale)	Mineurs disposant de leur propre électricité ; mineurs domestiques et de chauffage	Maintenant, borné par le halving et le hashprice	Réseaux basse tension, efficacité, chaîne d'approvisionnement en volume	Guerres des prix des acteurs établis, prix de l'électricité
Accélération de la preuve ZK	Réseaux de prouveurs, rollups	Attendre la convergence des systèmes de preuve	MSM et NTT programmables	La demande dépend des récompenses en tokens
Accélération FHE	IA chiffrée pour les entreprises et les États ; chaînes de confidentialité	2027–2029	NTT, systèmes à HBM, ingénierie cryptographique	Demande pas encore formée ; cycles de vente longs
Renforcement des TEE et inférence vérifiable	Réseaux d'inférence confidentielle, DePIN	Maintenant	Chiffrement mémoire avec intégrité, racines d'attestation indépendantes, vérifications ponctuelles ZK	Dépendance aux écosystèmes des fournisseurs
Chaîne ou réseau de calcul natif IA	Acheteurs d'inférence, mineurs, développeurs	Maintenant	Conception de la vérification, conception de l'émission, distribution	Déclin après le lancement, liquidité du token

Trois voies que je suivrais, chacune avec un premier pas

Si je devais choisir, je suivrais au plus trois voies, chacune avec un petit premier pas qui la teste avant d'engager de l'argent analyse.

- **Voie A : un accélérateur conçu d'abord pour l'inférence et capable de miner.** Premier pas : une étude d'architecture de quatre semaines qui modélise J/op et $/op pour un réseau GEMM INT8/FP8 résident en SRAM face à la RTX 5090 et au H100, à la fois sur le travail de Pearl et sur l'inférence de transformeurs, suivie d'une validation sur FPGA ou MPW en nœud mature avec des NRE de l'ordre du million de dollars.
- **Voie B : des services de silicium sur mesure avec la crypto comme premier client et l'IA comme second débouché.** Premier pas : transformer le modèle de NRE et de point mort de ce rapport en étude de faisabilité payante pour des projets de puces de calcul, avec des paiements par jalons et une recette sur une version figée du protocole.
- **Voie C : un réseau natif IA qui signe son acheteur de calcul avant son token.** Premier pas : une note de conception sur la vérification en couches (TEE plus échantillonnage plus vérifications ponctuelles ZK) et sur une arithmétique canonique neutre vis-à-vis des fournisseurs, rédigée avec un acheteur d'inférence nommément identifié avant toute conception de token.

CHAPITRE 12

12 JALONS
Ce qu'il faut surveiller, de novembre 2026 à 2027

Quand	Ce qu'il faut surveiller	Pourquoi c'est important
À tout moment	La hauteur d'activation FP8 de Pearl	Détermine si le matériel entier et les cartes non NVIDIA restent admis, et si la chaîne se scinde
Chaque mois	La part des blocs de Pearl portant du travail d'IA	Le meilleur test pour savoir si l'« utilité » devient réalité
Chaque mois	Les fournisseurs d'inférence répercutent-ils les revenus du minage sur leurs clients	La demande payante rencontre l'émission : le modèle dont Pearl a besoin
Chaque trimestre	Les revenus vérifiables de Bittensor face à ses émissions	Le plus grand réseau à travail utile comble-t-il son écart de subvention
Chaque trimestre	La part des flux x402 réellement initiés par des agents	Les paiements d'agents passent-ils de la signalisation au commerce
Janvier 2027	La hausse de prix de 3 % à 6 % de TSMC annoncée par la presse, surtout pour le 2 nm et le 3 nm	Le coût des tape-outs aux nœuds de pointe
Premier trimestre 2027	Le FP4 dans Pearl, selon l'AMA des fondateurs tel que rapporté	Un nouveau déplacement du travail vers les formats les plus récents de NVIDIA
Fin 2027	L'objectif d'équilibre de Samsung Foundry	Une seconde source moins chère reste-t-elle en bonne santé
Avril 2028	Le prochain halving du bitcoin	Divise par deux la cagnotte des mineurs SHA-256

CHAPITRE 13



QUESTIONS À POSER

13 Les questions que vous n'avez pas posées, mais auriez dû poser

Voici les questions que je poserais avant d'engager de l'argent ou du silicium dans un projet IA × crypto.

Qui contrôle la fonction de travail ? Si une petite équipe peut changer l'arithmétique à une hauteur codée en dur, la durée de vie de votre matériel dépend de sa décision. Demandez un engagement écrit de stabilité, ou concevez un matériel doté d'un second débouché.

Quel est le multiple de subvention ? Divisez les émissions (ou la valeur de marché) par les revenus payés depuis l'extérieur du réseau. Un multiple supérieur à 10 signifie que le réseau achète de l'activité au lieu de vendre un service.

Que coûte la prochaine étape, et quel chiffre vous arrête ? Un wafer multiprojet coûte des dizaines ou des centaines de milliers de dollars ; un jeu de masques coûte des millions. Fixez le chiffre d'arrêt avant de dépenser.

Quel est le hashrate d'amortissement, avec p ? Divisez les NRE par R × (1 − 1/k) × T × p. Si le résultat représente une grande part du réseau, la puce est un pari sur les règles, pas sur le silicium.

Quel est le second débouché de la puce ? Si le protocole fait un fork ou si le token chute, sur quoi d'autre le silicium peut-il se vendre ? Les GPU ont survécu au Merge d'Ethereum parce qu'ils en avaient un ; les ASIC mono-algorithme, non.

La valeur va-t-elle au token ou aux actions ? En 2026, les entreprises rentables de l'IA × crypto sont surtout des sociétés par actions dotées d'une distribution crypto ; décidez lequel des deux vous construisez avant de lever des fonds.

Qui est le premier client payant, nommément ? Chaque cas durable de ce rapport, de Grass aux sous-réseaux de calcul de Bittensor, a commencé par un acheteur nommé extérieur au réseau.

Ce sont des domaines que j'aimerais explorer avec d'autres, par l'échange technique et la collaboration technique : puces sur mesure pour le minage, l'entraînement et l'inférence d'IA, et accélérateurs ; développement d'IA ; et systèmes IA × crypto, de la chaîne au nœud et à la puce. Vous pouvez me joindre à i@tanshuai.com.

Limites de ce rapport

Il s'agit d'une recherche documentaire fondée sur des documents publics disponibles jusqu'au 13 novembre 2026. Je n'ai mené aucun entretien de première main ni aucune mesure indépendante pour cette édition. Plusieurs chiffres clés sont des déclarations d'entreprises ou de fournisseurs, comme les revenus d'Aethir, les reve-

nus de Grass, les performances d'Etched et de Taalas et l'efficacité de Cysic ; ils sont signalés comme tels. Les indicateurs de chaîne de Pearl proviennent d'un nœud complet indépendant et des pools, et changent d'heure en heure ; j'ai utilisé les relevés effectués vers 07:13 UTC le 13 novembre et je donne les fourchettes du jour. Les coûts de wafers, de masques et de conception sont des estimations de tiers et des valeurs de modèles qui ne concordent pas entre elles ; j'indique des fourchettes. Mon calcul du point mort des ASIC et le budget par étapes utilisent des hypothèses illustratives (k, T, p, NRE et coûts des étapes) que le lecteur devrait remplacer par les siennes.

À propos de l'auteur

Tan Shuai est technologue et fondateur, avec plus de vingt ans de travail couvrant les puces, le matériel et le logiciel système, et 20 brevets délivrés. Il a derrière lui de nombreuses années de R&D et de direction dans la conception de puces fabless, notamment la conception full-custom de chemins de données ASIC pour des puces de minage de cryptomonnaies et des puces de caméra multimédia, ce qui constitue l'angle de ce rapport : l'énergie par opération, le coût par wafer, ce qu'un protocole permet de figer dans le silicium, et qui paie. Son parcours l'a mené dans la Silicon Valley, à New York, à Chicago, à Taïwan, dans la région de la Grande Baie (Shenzhen et Hong Kong), au Vietnam et en Malaisie. Il travaille aujourd'hui comme chercheur et développeur en IA.

Échanges techniques : i@tanshuai.com · tanshuai.com

Avertissements

Ce rapport est une recherche documentaire indépendante fondée sur des documents publics disponibles jusqu'au 13 novembre 2026. Il est fourni à titre d'information et d'apprentissage uniquement et ne constitue ni un conseil en investissement ni un conseil d'ingénierie. Il ne traite que de technologie et d'affaires. Les prix, les indicateurs de chaîne et les règles des protocoles évoluent ; vérifiez auprès de la source primaire tout ce sur quoi vous vous appuyez. Les chiffres marqués selon le fournisseur, estimation ou analyse doivent être lus comme tels. Les prix des crypto-actifs sont très volatils.

© 2026 Tan Shuai. Vous pouvez partager ce rapport sans modification, avec mention de l'auteur.

Sources

Toutes consultées entre le 10 et le 13 novembre 2026. Regroupées par chapitre.

Pearl : protocole, code et données de la chaîne

- Komargodski et Weinstein, « Proofs of Useful Work from Arbitrary Matrix Multiplication », arXiv 2504.09971 : https://arxiv.org/abs/2504.09971
- Pearl Research Labs, livre blanc INT : https://pearlresearch.ai/research/int-whitepaper
- Code source de Pearl : bloc de genèse : https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/genesis.go ; paramètres de consensus et hauteurs des forks : https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/params.go ; comptabilisation du hash-rate : https://github.com/pearl-research-labs/pearl/blob/main/zk-pow/src/api/sanity_checks.rs ; guide de mise à niveau du fork à graine salée : https://github.com/pearl-research-labs/pearl/blob/main/docs/salted-seed-fork-upgrade-guide.md ; paramètres de la branche fp8-scheme : https://github.com/pearl-research-labs/pearl/blob/fp8-scheme/node/chaincfg/params.go
- Pearl Improvement Proposals : https://github.com/pearl-research-labs/pips ; PIP-3, minage en virgule flottante : https://github.com/pearl-research-labs/pips/blob/main/PIPS/pip-0003.md
- prlstats (nœud complet indépendant) : réseau : https://prlstats.com/api/network ; séries : https://prlstats.com/api/series ; accueil et modèle de revenus des mineurs : https://prlstats.com/api/home ; offre minée : https://prlstats.com/api/mined ; pools : https://prlstats.com/api/pools ; détenteurs : https://prlstats.com/api/holders ; hypothèses de matériel et de surcoût : https://prlstats.com/api/assumptions
- Kryptex Pool : réseau Pearl : https://pool.kryptex.com/prl/about-coin ; hashrate : https://pool.kryptex.com/prl/hashrate ; RTX 5090 : https://pool.kryptex.com/de/device/gpu/nvidia/nvidia-geforce-rtx-5090 ; RTX 4090 : https://pool.kryptex.com/device/gpu/nvidia/nvidia-geforce-rtx-4090 ; « Pearl FP8 fork » : https://pool.kryptex.com/articles/pearl-fp8-fork-en
- CoinGecko, Pearl : https://api.coingecko.com/api/v3/coins/pearl-2 ; https://www.coingecko.com/en/coins/pearl-research ; historique des prix : https://api.coingecko.com/api/v3/coins/pearl-2/market_chart?vs_currency=usd&days=365&interval=daily ; recherche du ticker « Perle » : https://api.coingecko.com/api/v3/search?query=perle
- SimeonOnSecurity, « Mining Pearl (PRL) in 2026 » : https://simeononsecurity.com/articles/pearl-prl-mining-guide-2026/
- Hashrate Index, « Pearl (PRL): Inside the AI-Compute Cryptocurrency », 2 juin 2026 : https://hashrateindex.com/blog/pearl-prl-ai-compute-cryptocurrency/
- Tom's Hardware, « New AI compute cryptocurrency Pearl sparks a GPU mining rush, but profitability is sliding » : https://www.tomshardware.com/tech-industry/cryptomining/new-ai-compute-cryptocurrency-pearl-sparks-a-gpu-mining-rush-but-profitability-is-sliding
- Together AI, partenariat avec Pearl Research Labs : https://together.ai/blog/together-ai-partners-with-pearl-research-labs ; page du modèle : https://together.ai/models/gemma-4-31b-it-pearl
- Basu, recensement des instances de minage de Pearl, arXiv 2606.04819 : https://arxiv.org/html/2606.04819v2
- Blockworks, lettre d'information 0xResearch, 28 octobre 2026 : https://blockworks.com/newsletter/0xresearch/issue/post_daf0d74a-2bf4-41e4-86d3-2a1275475a24
- Datawallet, « Pearl Explained » : https://www.datawallet.com/crypto/pearl-explained
- Coingabbar sur l'article de TheMarker consacré au financement de Pearl Research Labs : https://www.coingabbar.com/en/pearl-research-labs-funding-130m-stealth-round-themarker

Chaînes à travail utile et historique

- Grothe et al., analyse de sécurité de Gridcoin, USENIX WOOT 2017 : https://www.usenix.org/system/files/conference/woot17/woot17-paper-grothe.pdf
- Falsification des preuves d'apprentissage, arXiv 2208.03567 : https://arxiv.org/abs/2208.03567
- hoge.gg, vérification et confiance dans l'inférence décentralisée : https://hoge.gg/decentralized-inference-2026-verification-trust/ ; revenus et trésorerie de Bittensor : https://hoge.gg/bittensor-quantum-wallet-revenue-tao-treasury-2026/
- Kaspa : code matriciel de kHeavyHash : https://github.com/kaspanet/rusty-kaspa/blob/master/consensus/pow/src/matrix.rs ; lancement : https://coinranking.com/coin/V8GxkwWow+kaspa-kas/about ; annonce des IceRiver KS1/KS2 : https://www.iceriver.io/2023/03/14/1677644580/ ; histoire du passage des GPU aux ASIC : https://captainmining.com/blogs/crypto-news/the-evolution-of-kaspa-mining-from-gpus-to-asic-powerhouses ; hashrates des GPU : https://coincalculators.io/algorithms/kheavyhash ; KS7 : https://www.asicminervalue.com/miners/bitmain/antminer-ks7-40th ; réseau : https://minerstat.com/coin/kas ; 1 EH/s en 2024 : https://ourcryptotalk.com/news/kaspa-reaches-1-ehs-in-hashrate ; Crescendo : https://ourcryptotalk.com/news/kaspa-crescendo-live
- Aleo : FAQ du mainnet : https://aleo.org/post/aleo-mainnet-faq/ ; preuve décentralisée : https://aleo.org/post/decentralized-proving-advantages/ ; ASIC : https://whattomine.com/coins/432-aleo-zksnark/asics ; Blockworks, « State of Aleo Q3 2025 » : https://app.blockworks.com/report/state-of-aleo-q3-2025 ; ARC-46 : https://chaincatcher.com/en/article/2144326
- Qubic : The Block sur la réorganisation de Monero : https://www.theblock.co/post/366535/monero-faces-chain-reorganization-fears-after-qubic-says-it-controls-51-of-hashrate ; Halborn : https://www.halborn.com/blog/post/explained-the-monero-51-percent-attack-august-2025 ; minage de Dogecoin : https://hackernoon.com/qubic-is-now-mining-dogecoin-on-its-ai-compute-infrastructure ; CoinGecko : https://www.coingecko.com/en/coins/qubic

- Documentation du consensus de Bittensor : https://www.bittensor.com/llms.mdx/docs/internals/consensus/content.md ; FalconX, « State of Bittensor » : https://www.falconx.io/newsroom/state-of-bittensor-subnet-adoption-trends-network-mechanics-and-covenants-departure
- Primecoin : https://en.wikipedia.org/wiki/Primecoin ; Blockworks, « State of Filecoin Q3 2025 » : https://app.blockworks.com/report/state-of-filecoin-q3-2025
- Monero et forks de résistance aux ASIC : https://lbry.com/news/forking-for-asic-resistance ; fork de Sia : https://hotfix.coindesk.com/markets/2018/11/01/sia-completes-hard-fork-as-bricked-miners-form-rival-blockchain ; prix de l'Obelisk : https://www.cryptoglobe.com/latest/2018/11/obelisk-hardware-skyrockets-in-selling-price-after-sia-hard-fork/ ; les mineurs après le Merge : https://hotfix.coindesk.com/business/2022/11/11/ex-ethereum-miners-token-hop-to-stay-alive-after-the-merge

Secteurs IA × crypto

- DefiLlama : Render : https://defillama.com/protocol/render-network-bme ; Livepeer : https://defillama.com/protocol/livepeer
- Cointelegraph sur les revenus DePIN selon Messari : https://cointelegraph.com/news/depin-revenues-grow-despite-token-losses
- Aethir, bilan 2025 : https://aethir.com/blog-posts/aethirs-2025-wrap-up-decentralized-gpu-cloud-milestones
- CoinGecko, catégorie IA : https://www.coingecko.com/en/categories/artificial-intelligence
- Akash, prix de location du H100 : https://akash.network/the-bid/h100-rental-price-2026-cost-per-hour/ ; rapport du T1 2026 : https://akash.network/blog/akash-network-q1-2026-report/
- io.net, contrat de 8 millions de dollars et tokens quotidiens : https://lifestyle.kbew98country.com/story/271004/io-net-to-burn-up-to-12m-tokens-as-network-closes-8m-deal-and-hits-4-billion-daily-ai-tokens/
- DCD, contrat d'Axe Compute : https://www.datacenterdynamics.com/en/news/axe-compute-secures-15bn-compute-contract-with-unnamed-customer/
- Pluralis-8B : https://www.alphaxiv.org/abs/2607.13332.md ; Nous Research, Psyche : https://nousresearch.com/the-next-phase-of-psyche ; Implicator sur INTELLECT-3 : https://www.implicator.ai/prime-intellects-intellect-3-open-source-ambition-meets-centralized-reality/
- SiliconANGLE, financement de Prime Intellect : https://siliconangle.com/2026/07/08/prime-intellect-raises-130m-1b-valuation-ai-training-platform/ ; The Block, Venice : https://www.theblock.co/post/406934 ; RuntimeWire, financement de Nous : https://runtimewire.com/article/nous-research-75-million-round-1-5-billion-valuation
- ownyourmind, Gensyn : https://ownyourmind.ai/projects/gensyn/ ; Phemex, Ritual (titre seulement) : https://phemex.com/news/article/ritual-launches-ai-layer-1-mainnet-and-ritual-token-price-dives-72-from-high-101259
- Surcoût du calcul confidentiel, arXiv 2409.03992 : https://arxiv.org/abs/2409.03992v2 ; rapport 2025 de Phala : https://phala.com/posts/phala-2025-report ; TEE.fail : https://thehackernews.com/2025/10/new-teefail-side-channel-attack.html ; DeepProve, IACR ePrint 2026/1112 : https://eprint.iacr.org/2026/1112
- TRM Labs, mesure des paiements d'agents d'IA : https://www.trmlabs.com/trm-tech-blog/whos-actually-paying-measuring-ai-agent-payments-onchain ; Forkast sur la x402 Foundation : https://forkast.news/the-x402-foundation-activated-a-27-year-old-internet-code-200-million-transactions-later-the-real-volume-is-still-tiny/ ; The Block, mainnet de Tempo : https://www.theblock.co/amp/post/394131/tempo-mainnet-goes-live-with-machine-payments-protocol-for-agents ; BlockEden sur l'aGDP de Virtuals : https://blockeden.xyz/blog/2026/04/21/agdp-agent-gdp-virtuals-protocol-ai-blockchain-valuation-primitive-tvl-displacement/ ; The Block sur ElizaOS : https://theblock.co/post/410774/eliza-labs-native-token-dead
- Dealroom, revenus de Grass : https://dealroom.co/news/157794-grass-backed-by-multicoin-projects-75m-in-2026-data-revenue/ ; CryptoSlate, Fetch.ai et Ocean : https://cryptoslate.com/cryptos-flagship-ai-pact-fracture-fetch-sues-ocean-over-263m-fet-community-sales/ ; Fortune sur World : https://fortune.com/2026/06/01/internet-according-to-sam-altman-seeing-orb/
- Galaxy, capital-risque crypto : https://www.galaxy.com/insights/research/crypto-blockchain-venture-capital-q3 ; Crypto Briefing (titre seulement) : https://cryptobriefing.com/crypto-vc-investment-q3-2026-galaxy/ ; Blockhead sur Tiger Research : https://www.blockhead.co/2026/10/14/bitcoin-rose-43-in-q3-crypto-venture-funding-did-not-follow/ ; ForkLog sur les lancements de tokens de 2025 : https://forklog.com/en/85-of-tokens-launched-in-2025-fell-below-initial-prices/

Économie des ASIC et types de puces

- Hashrate Index, synthèse du 9 novembre 2026 : https://hashrateindex.com/blog/hashrate-index-roundup-november-9-2026/
- AnySilicon, coût de conception d'un ASIC par nœud en 2026 : https://anysilicon.com/news/how-much-does-asic-design-cost-in-2026-real-price-ranges-by-process-node/ ; Silicon Analysts, guide des coûts de tape-out pour start-up fabless : https://siliconanalysts.com/analysis/fabless-startup-tapeout-cost-guide ; données de marché : https://siliconanalysts.com/market-data
- MillionMiner, Bitmain S23 XP Hyd : https://millionminer.com/news/bitmain-wdms-2026-s23-xp-hyd-launch
- Khattak et Mikaitis, accumulation FP8 dans les cœurs tensoriels, arXiv 2512.07004 : https://arxiv.org/html/2512.07004v2
- Canaan, résultats du deuxième trimestre 2026 : https://finviz.com/news/389337/canaan-inc-reports-unaudited-second-quarter-2026-financial-results ; Bitdeer, résultats du troisième trimestre 2026 : https://greenstocknews.com/news/nasdaq/btdr/bitdeer-reports-unaudited-third-quarter-2026-results-record-revenue-of-us-320-7-million-adjusted-ebitda-up-173-to-us-69-6-million
- TheEnergyMag : Velaura AI : https://theenergymag.com/news/market-news/velaura-ai-series-a-bitcoin-pivot ; Core Scientific et Block : https://theenergymag.com/news/2026-07-29/core-scientific-exit-block-bitcoin-deal-ai ; économie des baux pour l'IA : https://theenergymag.com/news/2026-08-06/miner-weekly-ai-pumpability-out-fuel
- Blockworks, Cipher et Fluidstack : https://blockworks.co/news/cipher-mining-inks-3b-ai-hosting-deal-with-fluidstack ; Cointelegraph, Core Scientific : https://cointelegraph.com/news/core-scientific-to-quit-bitcoin-mining-by-year-end-as-ai-business-surges ; ForkLog, Bitfarms : https://forklog.com/en/former-bitfarms-shuts-down-all-mining-sites-in-the-us/ ; IREN, résultats du T1 de l'exercice 2027 : https://globenewswire.com/news-release/2026/11/05/3397020/0/en/iren-reports-q1-fy27-results.html
- Spark, concentration du marché des ASIC bitcoin : https://www.spark.money/research/bitcoin-asic-manufacturer-market-concentration

- Google, documentation du TPU7x : https://docs.cloud.google.com/tpu/docs/tpu7x?hl=es ; The Register, Ironwood : https://www.theregister.com/2025/04/10/googles_7thgen_ironwood_tpus_debut/ ; SemiAnalysis, B300 : https://inferencex.semianalysis.com/chips/b300 ; article sur le TPU v1, arXiv 1704.04760 : https://arxiv.org/abs/1704.04760 ; Glenn Lockwood, MI355X : https://glennklockwood.com/garden/processors/mi355x
- Tom's Hardware, Vera Rubin : https://www.tomshardware.com/pc-components/gpus/nvidia-launches-vera-rubin-nvl72-ai-supercomputer-at-ces-promises-up-to-5x-greater-inference-performance-and-10x-lower-cost-per-token-than-blackwell-coming-2h-2026 ; TrendForce, prix du CoWoS : https://trendforce.com/news/2026/10/27/news-tsmc-reportedly-eyes-20-cowos-price-hike-skips-equipment-price-renegotiations-amid-ai-demand-surge
- Menon Lab, comparaison des accélérateurs d'inférence d'IA : https://themenonlab.blog/blog/ai-inference-accelerators-compared ; Introl, les accélérateurs d'IA au-delà des GPU : https://introl.com/blog/ai-accelerators-beyond-gpus-tpu-trainium-gaudi-cerebras-2025
- StratNews (Reuters), puces d'IA nationales dans les centres de données financés par l'État : https://stratnewsglobal.com/team-sng/china-orders-state-funded-data-centres-to-use-domestic-ai-chips/ ; TNW, importations de H200 : https://thenextweb.com/news/nvidia-h200-chips-hong-kong-china-import-approval
- Succinct, preuve en temps réel sur 16 GPU : https://blog.succinct.xyz/real-time-proving-16-gpus/ ; produits ZK ASIC de Cysic : https://cysic-docs.pages.dev/hardware-products/zk-asic-products
- Tom's Hardware, Intel Heracles : https://www.tomshardware.com/tech-industry/cyber-security/intels-heracles-chip-computes-fully-encrypted-data-without-decrypting-it-chip-is-1-074-to-5-547-times-faster-than-a-24-core-intel-xeon-in-fhe-math-operations ; IEEE Spectrum : https://spectrum.ieee.org/fhe-intel ; Evertiq, Niobium et SemiFive : https://evertiq.com/design/2026-02-20-semifive-secures-design-win-with-niobium-for-fhe-accelerator ; The Quantum Insider, financement de Niobium : https://thequantuminsider.com/2025/12/03/niobium-23m-fhe-funding/ ; Optica OPN, Optalysys : https://www.optica-opn.org/home/industry/2026/february/optalysys_raises_us_$31_million_to_push_photonics_computing_toward_cloud_infrastructure/ ; Zama, HPU sur FPGA : https://www.zama.org/post/announcing-hpu-on-fpga-the-first-open-source-hardware-accelerator-for-fhe

Nœuds de gravure

- Silicon Analysts, prix des wafers : https://siliconanalysts.com/data/wafer-pricing
- Tom's Hardware, rumeurs sur les prix du TSMC A16 : https://www.tomshardware.com/tech-industry/semiconductors/tsmc-could-charge-up-to-usd45-000-for-1-6nm-wafers-rumors-allege-a-50-percent-increase-in-pricing-over-prior-gen-wafers ; Wccftech, prix de SMIC : https://wccftech.com/smic-5nm-chip-prices-up-to-50-percent-more-expensive/amp/
- EUROPRACTICE, calendrier MPW TSMC 2026 : https://europractice-ic.com/wp-content/uploads/2025/11/TSMC-Europractice-MPW-schedule-2026.pdf
- Horowitz, « Computing's Energy Problem », ISSCC 2014 : https://www.Gwern.net/doc/cs/hardware/2014-horowitz-2.pdf
- Tom's Hardware, feuille de route Huawei Ascend : https://www.tomshardware.com/tech-industry/artificial-intelligence/huawei-ascend-npu-roadmap-examined-company-targets-4-zettaflops-fp4-performance-by-2028-amid-manufacturing-constraints ; The Substrate, la chaîne d'approvisionnement chinoise en puces d'IA : https://www.the-substrate.net/p/where-chinas-ai-chip-supply-chain ; Tom's Hardware, Trainium3 : https://www.tomshardware.com/tech-industry/artificial-intelligence/amazon-launches-trainium3-ai-accelerator-competing-directly-against-blackwell-ultra-in-fp8-performance-new-trn3-gen2-ultraserver-takes-vertical-scaling-notes-from-nvidias-playbook

Opportunités et économie des lancements

- Webull (Benzinga), conférence sur les résultats de Broadcom : https://www.webull.com/news/15511095937917952 ; Quartr, Global Unichip T2 2026 : https://quartr.com/events/global-unichip-corp-3443-q2-2026_F9×6NJQY ; Taiwan News, Alchip : https://www.taiwannews.com.tw/news/6370946 ; Design & Reuse, VeriSilicon : https://www.design-reuse.com/news/202530457-surge-in-custom-chip-business-propels-verisilicon-s-q1-revenue-to-double-with-new-orders-skyrocketing-to-8-2-billion-yuan-increased-r-d-investment-leads-to-wider-net-loss-earnings-insight/ ; The Block, l'unité d'IA de Canaan : https://www.theblock.co/post/359250/crypto-miner-canaan-to-shutter-non-core-ai-semiconductor-business-unit-amid-strategic-realignment
- Clutch, tarifs du développement d'IA : https://clutch.co/developers/artificial-intelligence/pricing ; PixelPlex, coût du développement blockchain : https://pixelplex.io/blog/blockchain-development-cost/
- Hyperbolic, preuve d'échantillonnage : https://hyperbolic.ai/blog/deep-dive-into-hyperbolic-proof-of-sampling ; Gensyn, Verde : https://www.gensyn.ai/research/verde-a-verification-system-for-machine-learning-over-untrusted-nodes ; vérification en couches, arXiv 2512.20176 : https://arxiv.org/abs/2512.20176v1
- Sherlock, référence des prix d'audit pour 2026 : https://sherlock.xyz/post/smart-contract-audit-pricing-a-market-reference-for-2026
- Blockworks, critères de cotation de Binance : https://blockworks.com/news/binance-listing-standards-scrutiny ; Cointelegraph, accords de tenue de marché : https://cointelegraph.com/magazine/market-maker-deals-quietly-killing-crypto-projects ; Blockworks, Movement : https://blockworks.co/news/movement-rushi-manche-suspension-market-maker-web3port ; Nansen, HIP-3 : https://nansen.ai/post/what-is-hip-3-hyperliquid ; The Block, Monad : https://www.theblock.co/post/380094/monad-mainnet-launches

PREUVE DE TRAVAIL UTILE · CRYPTO × SILICIUM

Le calcul comme consensus

IA × crypto × puces en 2026, en partant de la preuve de travail utile de Pearl

Chaque chiffre clé de ce rapport indique sa source et sa date. Lorsque l'auteur calcule un chiffre, il est marqué comme une estimation ; lorsque seul le vendeur le communique, il est marqué comme une déclaration du fournisseur. Il ne traite que de technologie et d'affaires.

Tan Shuai

Technologue et fondateur · Chercheur et développeur en IA

tanshuai.com · i@tanshuai.com

ÉDITION PUBLIQUE · Données au 13 novembre 2026

