

מסמך מחקר · נובמבר 2026

חישוב כקונצנזוס

בינה מלאכותית × קריפטו × שבבים ב־2026, החל מ־Pearl

Pearl · עבודה מועילה · DePIN וסוכנים · כלכלת ASIC · צמתי תהליך · תקציבים מדורגים

תאן שואי

tanshuai.com · i@tanshuai.com

המשפט מחזיק; התמריץ לא

01 Pearl עובדת כהנדסה, עדיין לא כעבודה מועילה

כ־90 EH/s של כפל מטריצות במספרים שלמים מאבטחים את השרשרת, אבל רק כ־1% מהבלוקים נושאים עבודת AI אמיתית.

02 הביקוש בתשלום נמצא בהסקה, בנתונים ובפרטיות

ל־Bittensor כ־70 מיליון דולר בשנה של הכנסה ניתנת לאימות, ל־Grass 17 מיליון דולר במחצית הראשונה של 2026, TEE בייצור; תשלומי סוכנים כ־5,000–11,000 דולר בחודש.

03 ASIC משתלם רק עם קופה גדולה, כללים יציבים ושוק שני

תקרת הרווח שלו היא R × (1 - 1/k). כיום אף רשת אינה מציעה את שלושתם ניתוח.

≈90 EH/s

Pearl ב־13 בנובמבר 2026: בערך תפוקת המספרים השלמים של 122,000 מעבדי H100

≈1%

מהבלוקים של Pearl נושאים עבודת AI אמיתית (Blockworks)

פי 4–11

ההנפקה של Bittensor ביחס להכנסה הניתנת לאימות שלה

בינה מלאכותית × קריפטו בשישה תחומים: היכן הביקוש בתשלום אמיתי

01 רשתות עבודה מועילה

אמיתי אך קטן

≈70 מיליון בשנה

הכנסה חיצונית ניתנת לאימות של Bittensor

אות: הנפקה פי 4–11 מההכנסה; ≈1% Pearl בלוקי AI

המנגנונים עובדים; לתועלת דרושים משלמים

02 מחשוב מבוזר

אמיתי אך קטן

2.35 מיליון בשנה

עמלות on-chain של Render, הגבוהות ביותר שניתנות לאימות

אות: שווי שוק ≈פי 360 מהעמלות

בעיקר מכירה חוזרת של הסקה סחורתית

03 אימון והסקה

הרווח בהון

+100 מיליון בשנה

Prime Intellect, חברת הון על GPU מרוכזים

אות: רשת האימון של Gensyn אינה פעילה

ההתקדמות אמיתית; הרווח הולך להון

04 AI ניתן לאימות

מוקדם

398

משתמשים משלמים בענן ה־TEE של Phala, סוף 2025

אות: אין zkML בתשלום שפורסם

TEE בייצור; zkML מוקדם

05 תשלומי סוכנים

לפני ביקוש

5–11 אלף בחודש

תשלומי x402 שסוכנים יזמו בפועל

אות: ≈200 מיליון עסקאות, רובן איתות

המסילות הגיעו קודם; הביקוש לא

06 רשתות נתונים

קונים אמיתיים

17 מיליון

הכנסות נתונים של Grass, המחצית הראשונה של 2026 (דיווח עצמי)

אות: שווי שוק ≈פי 10 מההכנסה השנתית

אחד התחומים המעטים עם קונים אמיתיים

סבסוד הוא דלק להתנעה, לא מודל עסקי. הקריפטו מספק הפצה וסליקה; ההון לוקח את הרווחים, לפחות ב־2026.

הכפלה־צבירה אחת היא גיבוב אחד

ממשפט לרשת ראשית

Komargodski ו־Weinstein (אפריל 2025): הוכחת עבודה מועילה לכל מכפלת מטריצות בתקורה של 1 + o(1). הרשת הראשית של Pearl הושקה ב־27 באפריל 2026.

איך זה עובד

הכורים מכפילים מטריצות רועשות, מגבבים את הצובר של כל אריח לכרטיס הגרלה, ומוכיחים את האריח הזוכה בהוכחת Plonky2 של פחות מ־60 KB.

היחידה

H 1 הוא הכפלה־צבירה מזוכה אחת מסוג INT7 × INT7 → INT32: תקציב האבטחה של בלוקציין מתומחר בתפוקת טנזורים.

Pearl ב־13 בנובמבר 2026

89.77 EH/s

קצב הגיבוב של הרשת (צומת מלא בלתי תלוי)

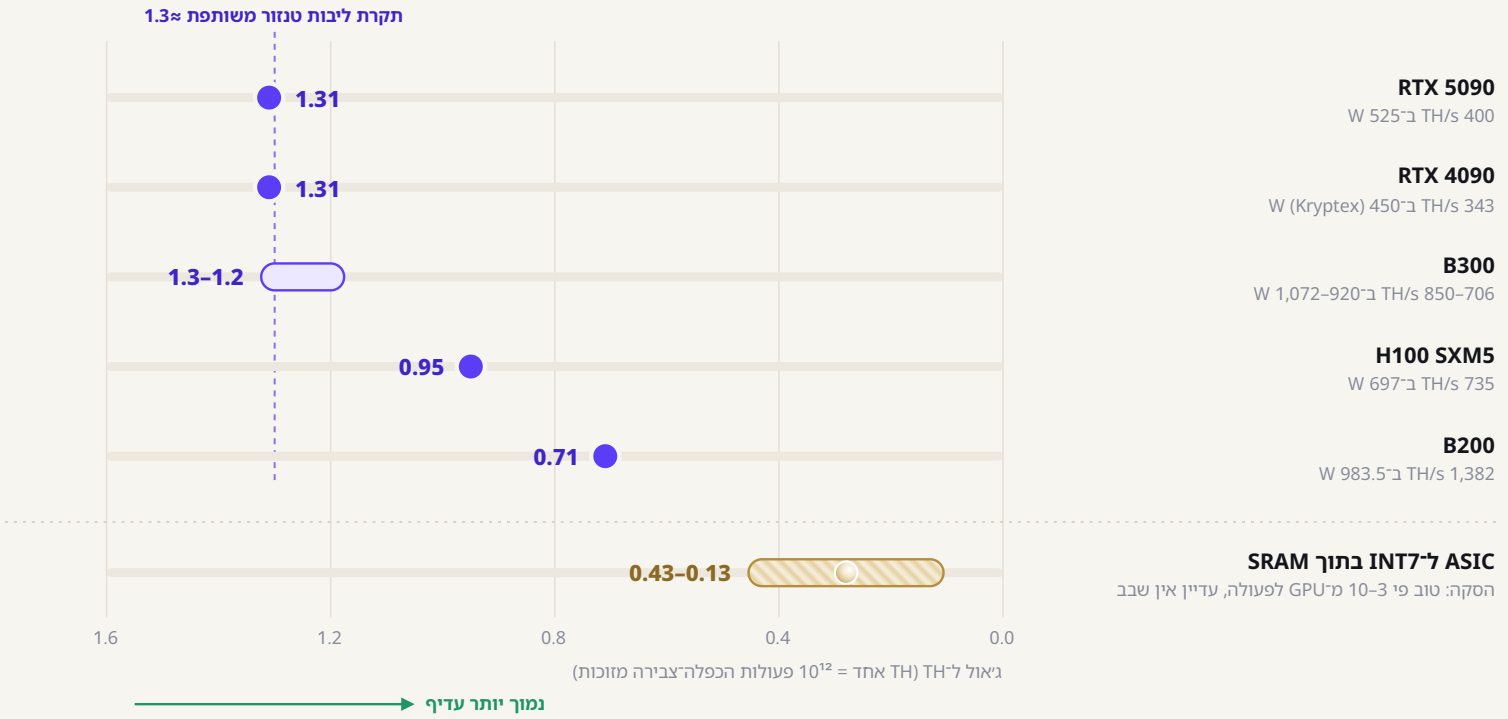
≈500 מיליון דולר

שווי שוק; בדילול מלא כ־2.8 מיליארד דולר

+78%

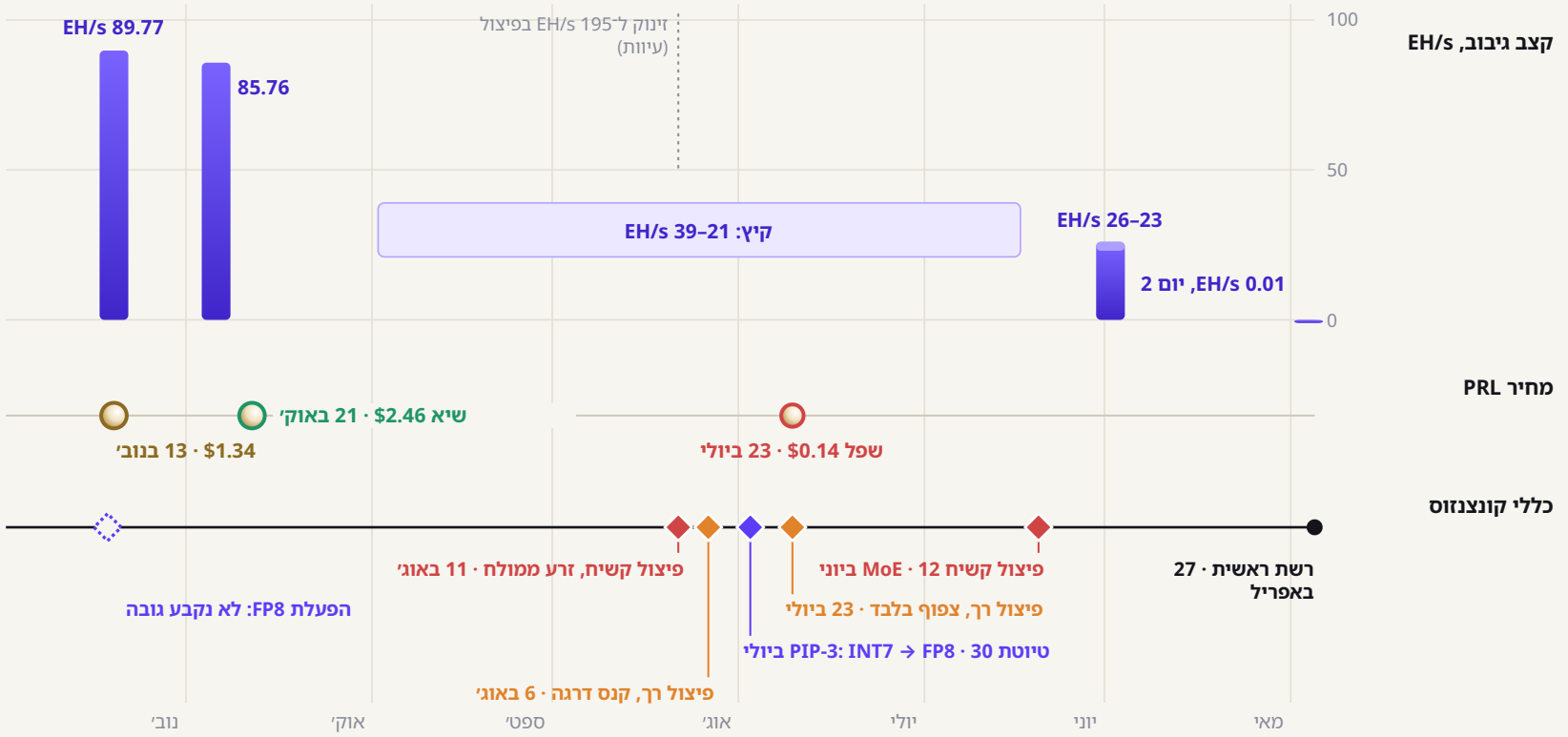
היצע חדש ב־12 החודשים הבאים

בעבודת המספרים השלמים של Pearl, ל־GPU תקרת אנרגיה משותפת



כרטיסי גיימינג וה־B300 נפגשים בכ־1.3 J/TH; אפילו ה־B200 מגיע רק ל־0.71. שבב שרק מכפיל מספרים שלמים של 7 סיביות מתוך SRAM יכול באופן סביר להיות טוב פי 3–10 ניתוח.

קצב הגיבוב עוקב אחרי המחיר; הכללים ממשיכים להשתנות



כ־120 מיליון מטבעות, 32% מההיצע של היום, נכרו בששת הימים הראשונים. RTX 5090 מרוויח כיום כ־5–6 דולר ביום; שכירת כרטיסים רק כדי לכרות מפסידה כסף.

המשפט מחזיק; התמריץ לא

כמה עולה עבודה מועילה

צד ההסקה: תקורה קטנה

Pearl מדווחת על 5.08% יותר זמן עבור Llama 70B על ארבעה H200; בדיקה בלתי תלויה מדדה תפוקה נמוכה ב־9.2% על GPU אחד מסוג Hopper.

כמה מרוויחה עבודה מועילה

צד הכרייה: מעט מדי קצב גיבוב

כרטיס שמשרת הסקה כורה בכ־27% מכורה ייעודי ב־prefill, ב־5.4% בצ׳אט, וכ־1% כשהפענוח שולט.

לכן כורים רציונליים מכפילים מטריצות אקראיות. התועלת היא בעיה מסחרית מחוץ לשרשרת: מישהו צריך לשלם על ההסקה.

ארבעה פיצולים בחודשיים, ו־FP8 בהמשך

12 ביוני
פיצול קשיח של תערובת מומחים (בלוק 71,935)

23 ביולי
פיצול רך למודלים צפופים בלבד (בלוק 91,630)

6 באוגוסט
פיצול רך של קנס דרגה: תצורות מסוימות מאבדות 50–87.5% מקצב הגיבוב האפקטיבי

11 באוגוסט
פיצול קשיח של זרע ממולח: כורים ישנים מייצרים רק שיתופים פסולים

לא נקבע
PIP-3: מ־INT7 ל־FP8, רק Hopper ו־Blackwell מתקבלים

2 מתוך 3
מועיל ל־AI, ניתן לאימות ביט־אחר־ביט, ניטרלי לחומרה: פורמט מספרים מקבל לכל היותר שניים

13 סיביות
נשמרות בצבירת FP8 של Hopper, עם קיטום: לא ניתן לשחזור בין דורות

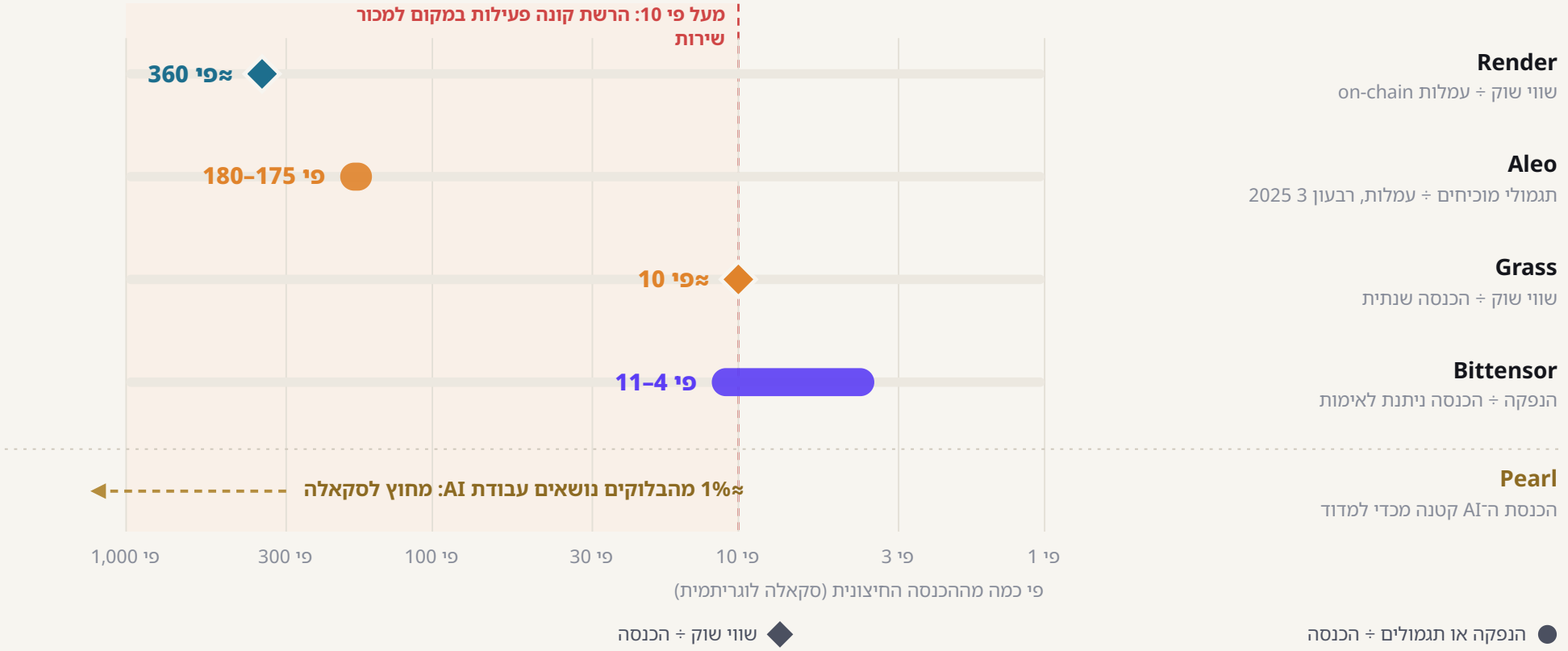
≈82%
מקצב הגיבוב על כרטיסים צרכניים ש־PIP-3 לא יקבל (מודל בביטחון נמוך)

מי ששולט בפונקציית העבודה שולט בגורלם של משקיעי החומרה ניתוח.

לא מטריצות ולא אפס ידיעה עוצרות ASIC

רשת	מה קרה	מה זה מלמד
Kaspa	מכפלת מטריצות 64 × 64 בין גיבובים; ASIC אחרי 16 חודשים, בערך פי 1,000 מיעילות GPU; הטוב ביותר מרוויח כיום 0.86 דולר ביום	מטריצה אינה מועילה ואינה חסינה ל־ASIC
Aleo	חידת SNARK סינתטית; מכונות ייעודיות אחרי חמישה חודשים; תגמולי מוכיחים פי 175–180 מהעמלות	עבודה סינתטית רצה על סבסוד
Qubic	טוענת לאימון AI; כוח החישוב שלה כרה Monero (לטענתה 51%) וכעת Dogecoin	PoW כללי יכול להילכד מבחוץ
Bittensor	ניקוד משוקלל לפי סטייק; כ־70 מיליון דולר בשנה של הכנסה ניתנת לאימות מול יותר מ־300 מיליון דולר של הנפקה	היחידה עם הכנסה חיצונית בהיקף משמעותי
Monero, Sia, Ethereum	פיצולים הפכו ASIC לחסרי תועלת; אחרי ה־Merge פחות מ־20% מהכורים מצאו עבודה חדשה	חומרה עם שוק שני שורדת

מכפיל הסבסוד הוא בדיקת הבריאות הכנה ביותר



מעל פי 10 בערך מההכנסה החיצונית, רשת קונה פעילות במקום למכור שירות. הכנסת ה־AI של Pearl קטנה מכדי להופיע על הסקאלה.

ביקוש אמיתי אך קטן, והערך עובר להון

שווקי מחשוב

עמלות on-chain של Render כ־2.35 מיליון דולר בשנה; H100 של Akash ב־2.73 דולר לשעה זול מ־AWS ומ־Google, אבל לא מ־Vast.ai או מ־RunPod.

אימון

8–40 מיליארד פרמטרים אומנו דרך האינטרנט; INTELLECT-3 חזר ל־512 H200 על InfiniBand.

AI ניתן לאימות

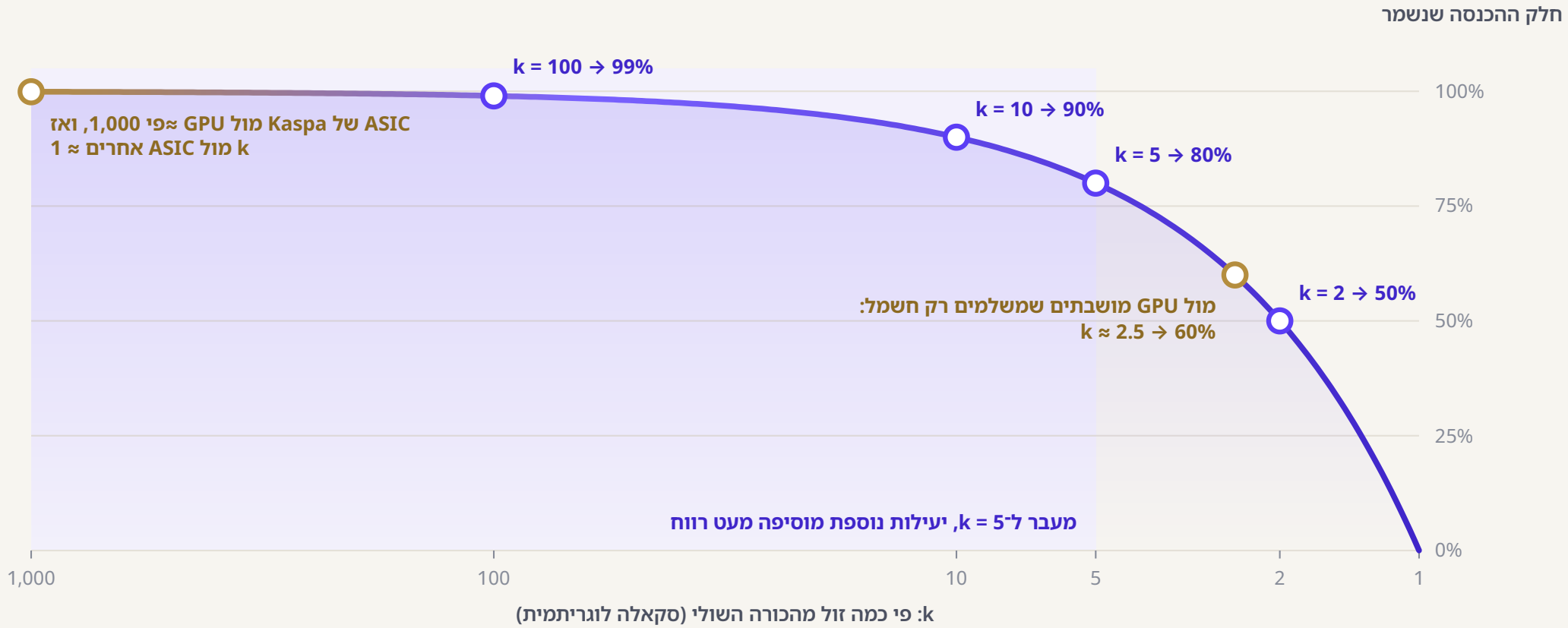
TEE מוסיפים פחות מ־7% ורצים בייצור; TEE.fail שבר את האישור (attestation) עם רכיב יירוט של 1,000 דולר; zkML מוכיח את GPT-2 ב־174 טוקנים לדקה.

סוכנים ונתונים

תזרימי x402 שסוכנים יזמו כ־5,000–11,000 דולר בחודש; ל־Grass 17 מיליון דולר הכנסות נתונים במחצית הראשונה של 2026.

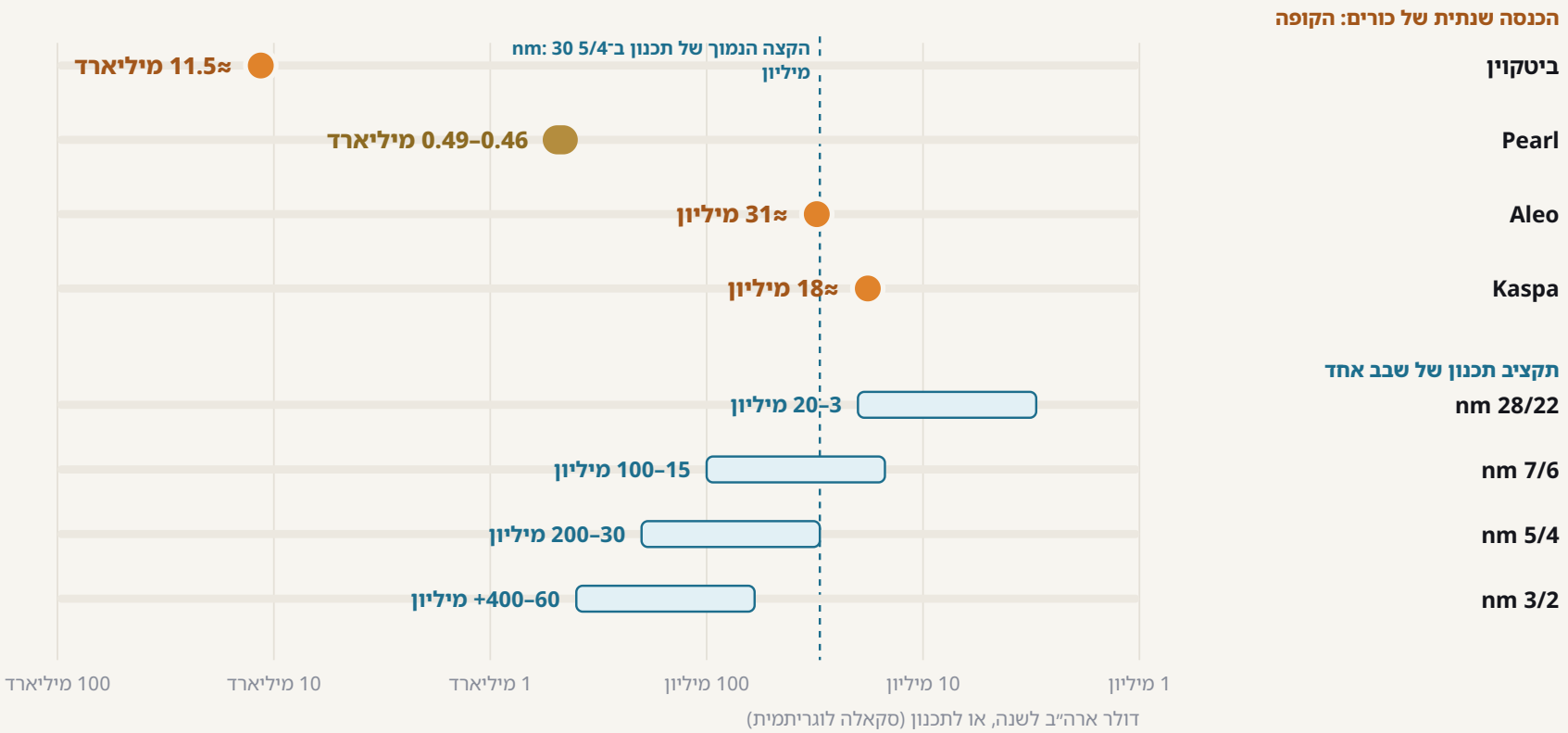
ארבע חברות הון על תשתית מרוכזת שוות כ־8.5 מיליארד דולר, בערך כמו כל טוקני ה־AI בשכבת הפרוטוקול ובשכבת היישומים.

מדוע יעילות מפסיקה להשתלם אחרי k = 5



תקרת רווח ≈ (1 - 1/k) × R. ברגע ש־ASIC הופכים לכורים השוליים, k מתאפס מול ASIC אחרים; מול GPU מושבתים שמשלמים רק על חשמל, k הוא כ־2.5. אומדן.

רוב הקופות קטנות מתקציב תכנון של שבב אחד



קצב הגיבוב להחזר [p × T × (1 - 1/k) × R] ÷ NRE = h*. עבור Pearl עם NRE של 20 מיליון דולר ו־k = 5: כ־4.9 EH/s למשך שנה אם הכללים מחזיקים, 9.8 EH/s אם הסיכוי שיחזיקו הוא חצי **אומדן**.

חמישה סוגי שבבים, חמש תשובות



שבב כרייה PoUW בסגנון Pearl

לא עכשיו

k: ללא HBM ויחידות כלליות: פי 3–10 לפעולה (הסקה)

שוק שני: הסקת AI, אם GEMM כללי ב־INT8/FP8

להמר על שבב הסקה שיכול לכרות, לא על רכיב INT7 ייעודי



שבב כרייה SHA-256 לביטקוין

להתרחק

k: ≤1 מול השחקנים הוותיקים

שוק שני: כמעט אין

רק עם חשמל עצמי, תצורת נישה, או כ־IP



שבב לאימון AI

IP, לא שבבים

k: פי 1.1–1.8 באותו דיוק

שוק שני: —

המארז וה־HBM מכריעים; לבנות IP, צ׳יפלטים, שירותים



שבב להסקת AI

לבנות

k: ויתור על כלליות: פי 10–20 לפי הטענות

שוק שני: זהו השוק העיקרי

לחתום עם הלקוח, ואז לבחור את הנעילה



מאיץ הוכחות ZK

להמתין ולגדר

k: פי 10–100 מ־GPU לפי הטענות

שוק שני: מוגבל מחוץ ל־ZK

מנועי MSM ו־NTT ניתנים לתכנות; FPGA קודם



מאיץ FHE

להתמקם מוקדם

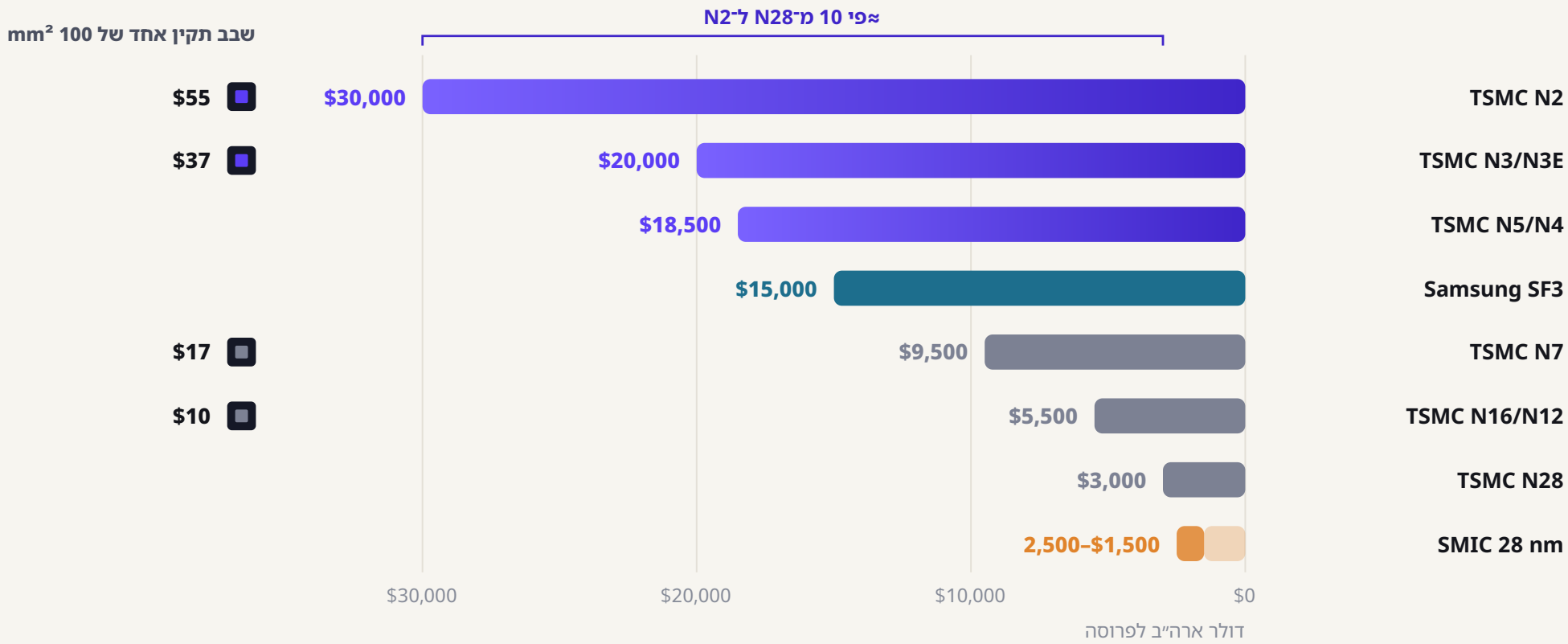
k: פי 1,074–5,547 מ־Xeon (Heracles)

שוק שני: חישוב מוצפן, AI פרטי

חלון 2027–2029; לאמת בנכסים קלים

אף רשת כיום אינה מציעה יחד קופה גדולה, כללים יציבים ו־k מעל 1: ה־k של ביטקוין שייך לשחקנים הוותיקים, הכללים של Pearl הם הפחות יציבים, והקופות של Kaspa ו־Aleo קטנות מדי.

פרוסה בחזית הטכנולוגיה עולה פי עשרה מפרוסה בצומת בוגר

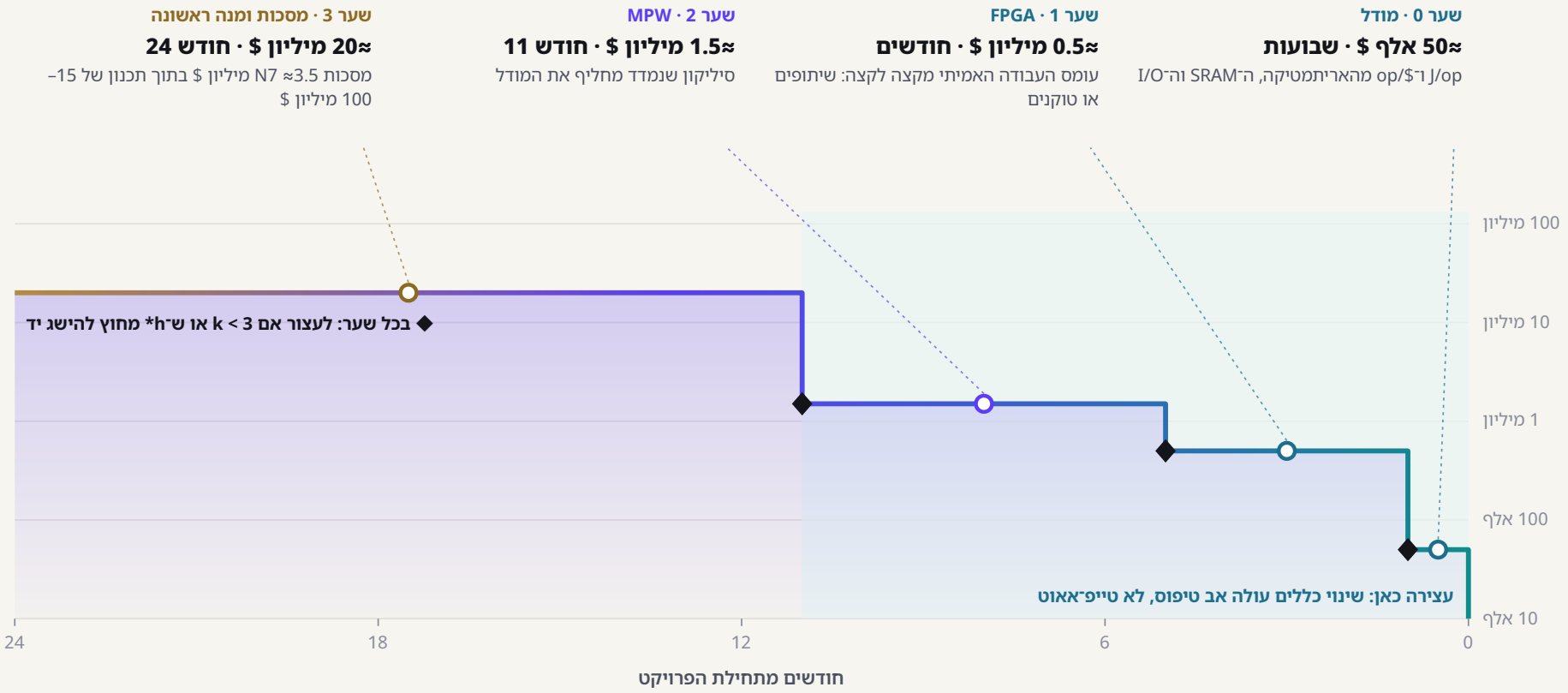


תנו לעלות הגדולה ביותר לאורך חיי השבב לבחור: אנרגיה לשבבי חישוב שפועלים מסביב לשעון, מארז ו־HBM לשבבי אימון (63% מעלות ה־H100), הבחירה בין SRAM ל־HBM לשבבי הסקה.

צומת לכל שבב ולכל צוות

שבב	צוות קטן	חברה בהיקף	מסלול סין היבשתית
שבב כרייה PoUW	MPW ב־28 או 16/12 nm; נפח ראשון N7/N6	N3/N3E או Samsung SF3	SMIC N+1/N+2 עובד, כ־40–50% יותר לשבב
שבב כרייה SHA-256	לא לרדוף אחרי החזית	דרגת 3 nm	קשה להשתוות ליעילות הצומת המוביל
שבב לאימון AI	IP, צ׳יפלטים, שירותי תכנון	N3P ואז N2/A16 עם CoWoS ו־HBM	SMIC N+2/N+3 עם HBM מקומי
שבב להסקת AI	ממוקד SRAM ב־N6/N7 או N4/N5/SF4	N3/N2 עם HBM, או צ׳יפלטים של SRAM	SMIC N+2 עם LPDDR או GDDR
מאיצי ZK ו־FHE	FPGA או MPW ב־16/12 nm קודם	N5/N3; FHE עם HBM	ZK ישים; FHE מוגבל בגלל HBM

לקנות ראיות לפני מסכות



מודל, FPGA ו־MPW מגיעים לסיליקון שנמדד בכ־1.5 מיליון דולר בתוך אחד־עשר חודשים; סט המסכות ומנת הייצור הראשונה עולים יותר מפי עשרה. לעצור לפני המסכות עד שהקונה או גרסת הפרוטוקול נקבעים.

הקריפטו כלקוח הראשון, ה־AI כשוק השני

שבבים מותאמים

- הביקוש הוא ברובו המכריע AI: ל־Broadcom כ־58 מיליארד דולר של מוליכים למחצה ל־AI בשנת הכספים 2026
- מיומנויות של שבבי כרייה עוברות: מערכים במתח נמוך, הספק של קילוואטים, טייפ־אאוטים מהירים
- לנצח עם ערוץ עצמאי לשוק, לא עם ערוצי הכרייה

פיתוח AI

- שעות עבודה הן סחורה: 25–49 דולר לשעה ב־Clutch
- הערך נמצא בחיבורים: קרנלים לשבבים מותאמים, כרייה בתוך שירות ההסקה
- הסקה ניתנת לאימות: TEE ועוד בדיקות מדגמיות ב־ZK

מערכות AI × קריפטו

- אימות בשכבות: TEE או דגימה, ואחר כך מחלוקות או בדיקות מדגמיות ב־ZK
- ההפצה והנזילות הן החסמים, לא העלות
- להחתים את קונה המחשוב לפני הטוקן

שלושה מסלולים, לכל אחד צעד ראשון קטן

מסלול	צעד ראשון
א. מאיץ שמכוון קודם כול להסקה ויכול גם לכרות	מחקר של ארבעה שבועות על J/op ו־$/op של מערך INT8/FP8 בתוך SRAM מול RTX 5090 ו־H100, ואחר כך FPGA או MPW בצומת בוגר
ב. שירותי סיליקון מותאם: קריפטו קודם, AI שני	להפוך את מודל ה־NRE ונקודת האיזון של הדוח הזה למחקרי היתכנות בתשלום עם אבני דרך וגרסת פרוטוקול קבועה
ג. רשת שנבנתה מלכתחילה ל־AI, עם קונה לפני טוקן	מסמך תכנון על אימות בשכבות ועל אריתמטיקה קנונית ניטרלית לספק, שנכתב עם קונה הסקה אחד בשמו

לכל היותר שלושה מסלולים, וכל אחד נבדק לפני שמתחייבים בכסף ניתוח.

על מה לעקוב, ושאלות שכדאי לשאול

ציוני דרך

בכל עת	גובה ההפעלה של FP8 ב־Pearl, והאם השרשרת מתפצלת
חודשי	חלק הבלוקים של Pearl שנושאים עבודת AI
רבעוני	ההכנסה הניתנת לאימות של Bittensor מול ההנפקה; תזרימי x402 שסוכנים יזמו
ינואר 2027	העלאת המחירים של TSMC ב־3–6% לפי הדיווחים, ל־2 nm ול־3 nm
רבעון 1 2027	FP4 ב־Pearl, כפי שדווח מה־AMA של המייסדים
אפריל 2028	החציה הבאה של ביטקוין

שאלות שכדאי לשאול

01 מי שולט בפונקציית העבודה?

02 מהו מכפיל הסבסוד?

03 כמה עולה השער הבא, ומה עוצר אתכם?

04 מהו קצב הגיבוב להחזר, עם p?

05 מהו השוק השני של השבב?

06 מיהו הלקוח המשלם הראשון, בשמו?

אלה תחומים שנוכל לחקור יחד, דרך חילופי ידע טכני ושיתוף פעולה טכני: שבבים מותאמים, פיתוח AI ומערכות AI × קריפטו.

i@tanshuai.com · tanshuai.com