

The Forgotten Engine

The Re-Rating of the Central Processing Unit

The AI hardware trade started with graphics chips. The central processing unit — cheaper to build, harder to replace, and increasingly the binding constraint in AI deployments — is the trade now underway.

INVESTMENT THESIS

The first act of the AI hardware cycle had one trade: own the graphics processing unit (GPU), and that call proved correct — the market has since priced it substantially. **The second act belongs to the central processing unit (CPU), and while several names have re-rated meaningfully, we believe the runway ahead is considerably longer than current prices imply.**

To understand why, it helps to think about what an AI agent actually does. It does not just generate text — it receives a task, breaks it into steps, and goes off to complete each one: searching the web, reading documents, running calculations, checking databases. The GPU handles the narrow moments of actual thinking. **Everything else — fetching information, routing results, keeping track of what has been done — falls to the CPU. In real-world agentic deployments, the GPU is actively working for as little as 10–12% of the time. The CPU is running for the other 88–90%.**

As AI moves from simply answering questions to autonomously completing tasks, the demand for CPU capacity per AI system grows continuously. Intel's management confirmed that the CPU-to-GPU ratio has shifted from roughly 1:8 during the model training era to 1:4 across today's inference deployments, with agentic workloads pushing the balance toward one-to-one. AMD raised its market size forecast for server CPUs to \$120 billion by 2030. NVIDIA — the world's largest GPU company — disclosed that it expects nearly \$20 billion in revenue this year from its own new CPU product alone, entering a \$200 billion market it had never previously addressed. These are not analyst projections — they are the companies actually building this infrastructure describing what they see in front of them.

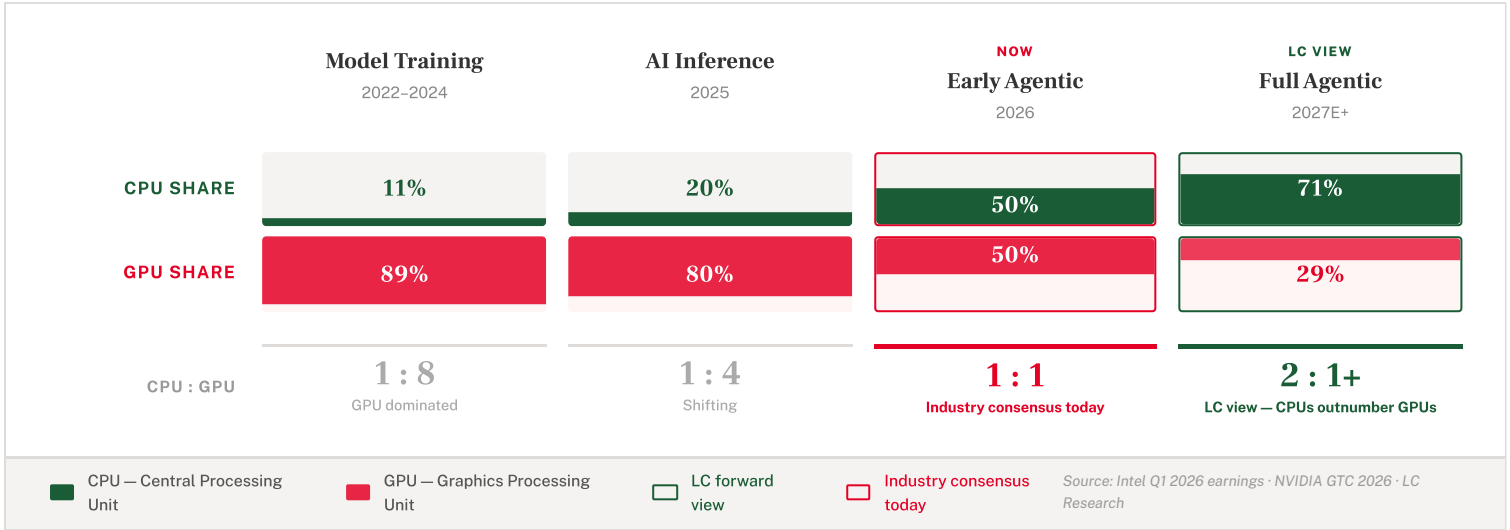
The industry tends to view one CPU per GPU as the natural endpoint. We think that underestimates where this goes. The cost gap between GPUs and CPUs is substantial — GPUs carry a significant manufacturing premium because of the specialised High Bandwidth Memory (HBM) packaging bonded directly to the chip, a component that accounts for a large share of production cost and is not required in CPUs. When an operator's GPU sits idle for most of the working day waiting for the CPU to finish its coordination work, adding more CPUs is not just logical — it is inexpensive relative to the asset being protected. **That arithmetic does not resolve at parity — it keeps moving in the CPU's favour.**

FIVE WAYS TO OWN THE TRADE

COMPANY	INVESTMENT THESIS	LATEST DEVELOPMENTS
AMD PREFERRED PURE PLAY	The only company selling CPU, GPU and networking as a single AI rack platform — with no direct x86 competition until mid-2027. Re-rating depends on customers buying the full stack, not the GPU alone.	Q1 2026: rev \$10.3B +38%; DC \$5.8B +57% · EPS \$1.37 vs \$1.27E · Q2 guide \$11.2B · Server CPU TAM \$120B+ by 2030
Intel PREFERRED PURE PLAY	Unable to meet its own demand, and chosen by NVIDIA for its flagship AI system despite NVIDIA having an in-house alternative. Foundry needs an external customer to re-rate fully.	Q1 2026: DCAI \$5.1B +22%; op. margin 31% · Q2 DCAI up double digits QoQ · 18A yields 2 qtrs ahead · Xeon 6+ (288 cores) launched Computex June 2026
NVIDIA DIVERSIFIED BENEFICIARY	NVIDIA built a CPU to stop its own GPUs sitting idle — the most powerful confirmation this bottleneck is real. US export controls removing China from guidance are the main near-term overhang.	Q1 FY2027: rev \$82B +85%; DC \$75B +92% · Q2 guide \$91B (excl. China) · ~\$20B Vera CPU revenue FY2027 · Vera Rubin NVL72 ships Q3 2026
Arm Holdings NEW ENTRANT	Inside virtually every major AI chip being built today. Launching its own chip shifts Arm from royalties to full chip economics — demand doubled within weeks of launch. Competing with NVIDIA and Amazon puts the royalty base at risk.	FY2026: \$4.92B +23% · DC royalties 2x+ YoY · Q1 FY2027 guide \$1.26B +20% · AGI CPU demand \$2B+ (FY2027–28)
Qualcomm NEW ENTRANT	Priced as a fading smartphone business — yet has a confirmed AI chip order from a major cloud company shipping this year, valued at zero by the market. Apple's modem exit and an unproven data centre business are the two risks.	Q2 FY2026: rev \$10.6B; EPS \$2.65 · Auto \$1.3B +38% — run rate above \$5B · Hyperscaler DC shipment Dec 2026 · Q3 guide \$9.2–10.0B
MediaTek NEW ENTRANT	The world's largest smartphone chip company — with manufacturing scale, cost structure and Asian cloud relationships most rivals lack. Data centre CPU ambition is early but the industrial logic is compelling.	No DC CPU product yet · Dimensity edge AI chips in production · Taiwan-listed (2454.TW) · ~15x forward P/E · AI partnerships with Google, Microsoft, NVIDIA

THE RATIO SHIFT — HOW CPU AND GPU BALANCE IS CHANGING

CPUs cost a fraction of GPUs — partly because they do not require the expensive High Bandwidth Memory (HBM) packaging bonded directly to the chip that makes GPUs so costly to manufacture. That cost gap makes provisioning more CPUs economically compelling. The industry estimates a 1:1 ratio as the endpoint. We think it goes further.



THE CASE FOR MORE RUNWAY

01

THE GPU IS IDLE MOST OF THE TIME

In the era of training large AI models, one CPU could support seven or eight GPUs. That ratio made CPU demand look irrelevant against the GPU boom. It no longer holds. Intel confirmed on its Q1 2026 earnings call that in today's AI deployments, the ratio has already moved to one CPU per four GPUs — and in *agentic* deployments (where AI completes tasks autonomously), it is approaching one-to-one. More strikingly, Intel's own customer data shows that GPUs in these systems are actively working only 20–30% of the time. **The GPU, the most expensive component in the rack, sits idle for the majority of every working day.** Server CPU prices have risen 10–20% since March 2026 as the market begins to price this in.

02

ELECTRICITY IS THE NEW CONSTRAINT

Running an AI data centre costs an enormous amount of electricity — and power capacity has become a genuine constraint on how fast the industry can grow. Arm projects that next-generation AI workloads will need **four times as many CPUs per megawatt** of electricity versus today's systems. CPUs now represent **44% of total power consumption** at this scale, making energy efficiency the decisive battleground. Arm-based chips win that battle, which is why they now power **roughly half of all deployments at the world's largest cloud providers.**

03

BETTER SOFTWARE IS MAKING THE PROBLEM BIGGER

Better software normally means you need less hardware, but in AI the opposite is happening. A recent update to Python — the coding language most AI systems are built on — unlocked the ability to run **four times as many tasks at once.** More tasks simultaneously means more demand on CPUs, not less. Add to that the rise of multi-agent systems, where dozens of AI agents work in parallel on different parts of the same problem, and the multiplier effect becomes significant. Qualcomm's CEO summarised the trajectory at Computex 2026: the volume of AI processing the world demands every ten seconds will be **forty times larger by 2030.**

A BOTTLENECK THAT CANNOT BE SOLVED BY BETTER SOFTWARE

When GPU chips became scarce and expensive, engineers responded brilliantly — they found ways to make the same hardware do ten times more work by compressing models and cutting computational corners. This is why a chip shortage that looked like it would last years was partly resolved in months. Investors have rightly come to expect that AI hardware constraints get solved.

Every previous constraint in AI hardware was eventually solved by making software more efficient — doing the same work with less compute. That will not happen here. The CPU's job in an agentic system is not approximate. A database must return the right result. An API call must complete. A web page must load. **You cannot compress or shortcut any of those tasks. The bottleneck that cannot be engineered away is the one worth owning.**

ISA (INSTRUCTION SET ARCHITECTURE) — THE ARCHITECTURE THAT DETERMINES WHO GETS PAID

Not every company benefits equally from rising CPU demand. The key is something called the Instruction Set Architecture (ISA) — essentially the language a chip speaks. It determines who can build competing chips, who gets paid when demand rises, and how much they get paid. There are three architectures that matter in this story.

X86 Intel & AMD THE ENTERPRISE MOAT Only Intel and AMD hold x86 licences. Decades of enterprise software certifications and switching costs constitute a moat that technical performance alone cannot displace. When CPU demand rises, Intel and AMD capture the premium directly at the socket level — server CPU prices already up 10–20% since March 2026. MOAT TYPE Enterprise software lock-in WEAKNESS Less efficient per unit of electricity WHO PROFITS Intel & AMD capture it fully	ARM Arm Holdings WINNING ON EFFICIENCY — TRANSITIONING ON ECONOMICS Arm designs chip blueprints and licenses them to others — Apple, NVIDIA and Amazon all build on Arm's architecture. Its chips use less electricity than x86, which matters enormously when data centres are fighting over power capacity. Arm has recently launched its own chips for sale, shifting from collecting a small royalty on others' sales toward earning the full selling price — a transition that significantly changes the economics of this business. MOAT TYPE Best performance per unit of electricity EVOLUTION Transitioning from royalties to chip sales TENSION Competing with its own paying customers	RISC-V Open Source A SIGNAL, NOT YET A POSITION RISC-V carries no licence fee. The ecosystem is not yet competitive in production data centres, but strategic investment activity is accelerating. NVIDIA has backed SiFive and opened NVLink Fusion to RISC-V CPUs. Qualcomm acquired Ventana Micro Systems in December 2025 — a RISC-V specialist — principally to deepen its Oryon CPU team. These are hedges: if ARM's own chip ambitions create licensing tensions, RISC-V becomes the alternative of first resort. MOAT TYPE No fees, fully customisable WEAKNESS Not yet ready for production at scale SIGNAL NVDA & QCOM have invested as a hedge
---	--	--

RISKS TO THE THESIS

WHAT COULD GO WRONG	WHY IT MATTERS
Software gets more efficient	Better scheduling and memory management will reduce CPU demand at the margin, but cannot eliminate the core requirement — every external task an AI agent completes needs a CPU to handle it. AMD and NVIDIA are both building their businesses on the assumption that this bottleneck persists.
The big cloud companies build their own	Amazon, Google and Microsoft are all designing their own server chips rather than buying from Intel or AMD. If this deepens, it narrows the addressable market for x86 incumbents — though Arm collects a royalty regardless of who manufactures the chip.
AI agents take longer to scale than expected	Most enterprises are still using AI as an assistant rather than an autonomous agent. If that transition takes one to two years longer than expected, the CPU demand inflection is real but deferred. Intel and AMD's latest earnings suggest it is already underway.
Intel's factory story needs proof	Intel's chip business is delivering — Xeon 6 Plus launched on 18A in June 2026 with yields ahead of plan. The foundry-for-others business still needs a named external customer. Without one by year-end 2026, the foundry premium in Intel's valuation compresses.
Arm's customers start building around it	By launching its own chips, Arm now competes with the customers paying its royalties — NVIDIA, Qualcomm and Amazon. Qualcomm's acquisition of a RISC-V specialist in December 2025 is an early signal that alternatives are being explored.
Export restrictions cut deeper	NVIDIA has already excluded China from Q2 2026 guidance entirely. Huawei's Ascend 950PR entered mass production in March 2026 and is filling the gap domestically. Further tightening could extend the impact across the coverage universe.

LIGHTHOUSE CANTON

Singapore

16 Collyer Quay #11-02
Collyer Quay Centre
Singapore 049318

☎ +65 67130570

India

Unit No-104A, Worldmark
2 Asset Delhi Aerocity,
New Delhi 110037

☎ +91 9650473961

Unit No 507/508, A Wing,
INS Tower, G Block, BKC,
Mumbai- 400051

☎ +91 9650473961

UAE

The Exchange Gate Village 11,
Unit 802 Dubai International
Financial Centre PO Box 507026
Dubai, UAE

☎ +971 45 861500

1st Floor, WeWork37,
Cunningham Cross Rd,
SRT Road Vasant Nagar,
Bengaluru-560001

☎ +91 9900096873

RK Swamy Centre, Hansa
Building, Door No:3, Thousand
Lights,
Chennai-600006

☎ +91 9650473961

UK

24 Hanover Square,
London W1S 1JD

☎ +44 164 2843 487

Suite 502, Building 450, Central
Plaza, Genome Valley,
Shameerpet,
Hyderabad 500 078

☎ +91 9650473961

Unit No FF-10, FF Floor, Pragya
Accelerator, Block 15T GIFT CITY,
Gandhinagar
Gujarat-382355

☎ +91 9650473961

✉ info@lighthouse-canton.com

📌 Lighthouse Canton

DISCLAIMER

The contents of this document are confidential and are meant for the intended recipient only. If you are not the intended recipient, please delete all copies of this document and notify the sender immediately.

This document, provided as a general commentary, is for informational purposes only and is not to be construed as an offer to sell or solicit an offer to buy any financial instruments in any jurisdiction. This does not constitute any form of regulated financial advice, and your independent financial advisor should be consulted prior to taking any investment decision(s).

This document is based on information from Sources which are reliable but has not been independently verified by Lighthouse Canton Pte Ltd and its affiliate companies ("LC"). LC has taken the reasonable steps to verify the contents of this document and accept no liability for any loss arising from the use of any information contained herein. Please also note that past performances are not indicative of future performance.

Information contained herein are those of the author(s) and does not represent the views held by other parties. LC is also under no obligation to update you on any changes made to this document.

This document is prepared by Lighthouse Canton Pte Ltd and its affiliate company, Lighthouse Canton Capital (DIFC) Pte Ltd, which are regulated by Monetary Authority of Singapore ("MAS") and Dubai Financial Services Authority ("DFSA") respectively. MAS and DFSA has no responsibility for reviewing, verifying and approving the contents of this document and/or other associated documents. The contents of this document may not be reproduced or referenced, either in part or in full, without prior written permission from LC.

This document is confidential and is only intended for Accredited Investors and/or Professional Clients, as defined by MAS and DFSA.