Pure-play datacenter AI silicon is delivering exits
The next category is silicon for physical AI

Babak Soltanian, Ph.D.
CEO and Co-founder, Tayen

In mid-2025, the consensus was that AI silicon was finished and Nvidia had locked it with AMD a distant second, but the past two quarters proved otherwise. In December 2025, Nvidia acquired Groq's IP, leadership, and engineering core for $20 billion, a generational outcome for a young pure-play chip startup.

In May 2026, Cerebras opened on Nasdaq at $385, more than double its $185 IPO price and the first major pure-play AI silicon listing in a decade. These exits alone mark a category arrival and the leading edge of a deeper pipeline.

Four things I think are now true:

1. **Capital efficiency and moat**

   The capital required to build a frontier AI silicon company is similar to what frontier model and humanoid startups now raise, but the moat is deeper. A trained world model can be replicated by the next one while a working AI chip cannot.

2. **Premium outcomes**

   Groq sold in three months at 2.9x its last private round. Cerebras priced above range and doubled on day one. The "GPUs are enough" debate is over.

3. **The exit ladder is rebuilt**

   Strategic M&A at $20B+. Public IPOs at $50B+ entry. The standard counterargument against funding chip startups (no credible exits under $1B in revenue) no longer holds.

4. **Architectural conviction beats caution**

   Both winners built architectures that looked exotic at seed. The conviction premium in silicon is now publicly priced, and this conviction is more prominent for physical AI silicon, where heterogeneous and highly optimized architectures are the entry ticket.

Bigger market, more winners ahead

The multi-billion-dollar outcomes happened alongside Nvidia's continued market-cap expansion in the same window. The pie got larger and made room for multiple winners with different products. That dynamic does not weaken when workloads move off the rack, but it accelerates.

The next category is being funded

Datacenter AI silicon was the first category, and its main customers (hyperscalers) built out their own architecture roadmaps long before the chip startups arrived. The architecture caught up to demand that was already obvious.

The next category is inverted. The customers are already funded, but the silicon is not there yet. Over the past eighteen months, well-capitalized companies have been building systems that need on-device frontier intelligence:

- **World models.** Physical Intelligence ($600M at $5.6B), AMI Labs ($1B at $4.5B), Skild AI ($1.4B at $14B), World Labs ($1B at $5B), Wayve, Waabi. Billions in aggregate across companies building world models for physical systems.
- **Humanoids.** Figure ($2B+ raised at $39B Series C), 1X, Apptronik, Agility, Sanctuary. Companies whose unit economics require on-device autonomy at scale, not cloud-tethered demos.
- **Robotics and autonomy.** A second tier of mobile robotics, drone autonomy, and untethered industrial systems. All requiring inference inside the loop, every cycle.

These companies are funded, hiring, and deploying. They share one constraint: the silicon to run frontier world models on these devices, inside their power and latency limits, doesn't exist yet. It is just emerging as a new product category. Datacenter AI silicon caught up to its customers. The next category needs to catch up to customers who are already at scale or actively scaling.

Why this category is different by design

At Tayen, we believe the constraints distinguishing physical AI from datacenter silicon are not incremental. They are orthogonal:

- **Latency in milliseconds.** Sensor-to-actuator paths that close inside the device, every cycle. Missing the budget means the task fails.
- **Power one to two orders of magnitude below datacenter.** Tens of watts, often single-digit. Wafer-scale and kilowatt-class GPU economics break down before this threshold.
- **Heterogeneous workloads.** Frontier world models combine autoregressive prediction, diffusion sampling, geometric reasoning, simulation, and closed-loop control. Each compute signature is distinct. Homogeneous substrates forfeit more than they gain.
- **Volume that dwarfs the datacenter.** By the end of the decade, untethered AI-bearing devices will outnumber datacenter GPUs by orders of magnitude. Unit economics will be set by humanoid OEMs, vehicle platforms, and robotics primes, not hyperscalers.

These constraints sit outside Nvidia's economic engine. The incumbent has neither the incentive nor the architecture to compete for this segment as a core business. That is what leaves the category open.

Three predictions

1. **A Cambrian period of AI silicon**

   Groq, Cerebras, and further pure-play exits likely in the coming quarters will accelerate capital formation across AI silicon, not just for datacenter plays. Funds that previously avoided semis as "structurally unfundable" will allocate. Funds that participated in this cycle's winners will look for the next category.

2. **Physical AI silicon emerges as a dominant category**

   It moves from emerging to actively priced over the next four-to-six quarters. The seed and Series A rounds priced in this window will define the next decade of AI hardware.

3. **Design velocity becomes existential**

   Cerebras took a decade between WSE-1 and where it is today. That cadence is no longer competitive against the pace at which frontier models evolve. The next generation of AI silicon companies will need design methodologies that compress chip development cycles from years to months. Companies that can't will fall behind.

Where Tayen fits

At Tayen, we build Frontier Silicon: a heterogeneous multi-engine AI processor for world models running on-device, under hard power and latency budgets. We build for the category described above. Not the datacenter, not Nvidia's next move, but the silicon physical AI's already-funded customers need.

Our central metric is TTL (Time-to-Learning), the elapsed time of one full sensorimotor update cycle. Miss it and the task fails; there is no graceful degradation. We wrote up the framework in a separate whitepaper.

We build on a proprietary AI-native autoregressive silicon methodology that compresses design cycles from years to months. That is the iteration-velocity bet underneath the architectural bet.

The investors who priced Cerebras's seed in 2016 backed an against-the-stream thesis; kudos to them. The market made room for that bet alongside Nvidia, not at its expense. A decade later, their bet paid off. I believe we are in the analogous moment for physical AI silicon: a category open to new entrants, at the start of a wave whose customers are already in place.

About Tayen

Tayen's three Ph.D. co-founders bring 50+ combined years across Silicon Valley startups and tech bellwethers, 100+ prior tapeouts, 30+ patents, and 250+ publications across AI, silicon, autonomy, and wireless systems. The cross-disciplinary depth required to deliver Frontier Silicon is rare, and our team has it. Our advisory board extends that depth across the same domains.