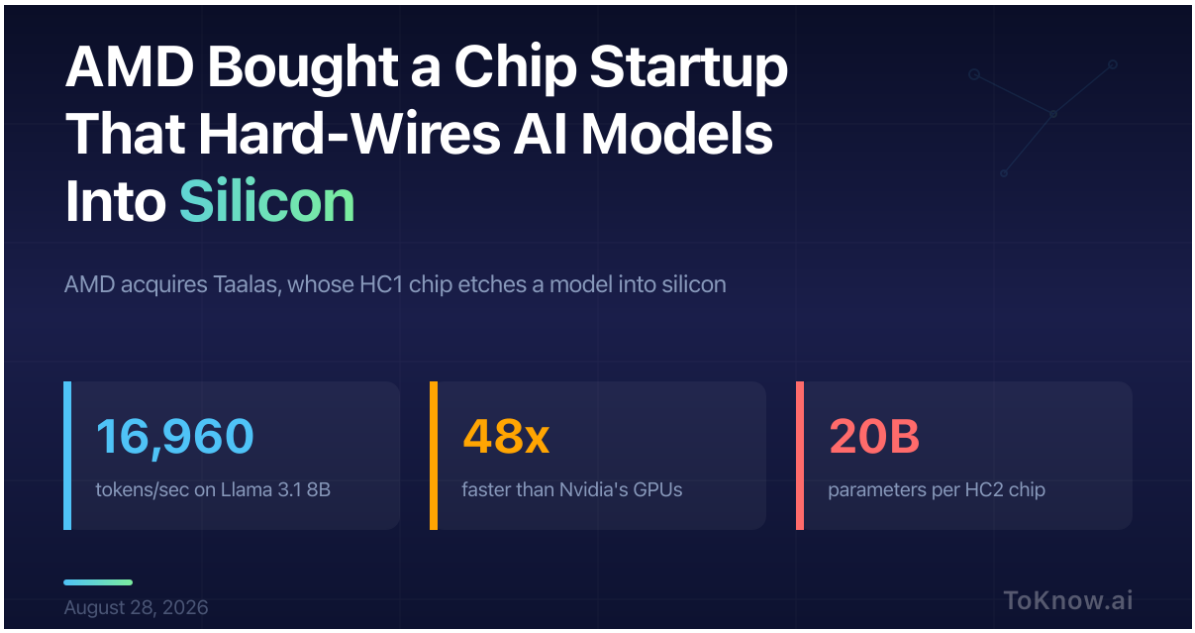
AMD Bought a Chip Startup That Hard-Wires AI Models Into Silicon

Kabui, Charles

2026-08-28

Read at ToKnow.ai



AMD agreed on August 6, 2026 to acquire Taalas, a Toronto startup that hard-wires a finished AI model into the chip itself. Its HC1 processor, first shown in February, is built on TSMC's 6-nanometre process and stores a model's weights directly in its transistors, with on-chip SRAM for the temporary state the model keeps while generating text. Because the weights live in the logic rather than external memory, there is no DRAM read on every forward pass. In Taalas'

own benchmarks, a single HC1 served Meta's Llama 3.1 8B at 16,960 tokens per second, 48 times faster than Nvidia's GPUs and 8.5 times faster than Cerebras' accelerators. AMD will pair it with its Instinct GPUs and Helios racks, with the deal expected to close in the fourth quarter.

A GPU runs any model; an HC chip is stuck with the one etched into it, and anything larger than a fine-tuning adapter needs new silicon, though Taalas says just two metal layers change, at roughly 100 times less cost than training a frontier model. For providers running stable, high-volume models, that buys huge speed and energy savings: second-gen HC2 chips at 20 billion parameters each would serve a trillion-parameter model with about 50 accelerators. The deal follows Nvidia's $20 billion purchase of Groq assets, another sign that inference-specific silicon is becoming a real second track.

Hardware is starting to be chosen per model rather than per workload. AMD still expects GPUs to hold most of the AI chip market, but hard-coded silicon becomes attractive once a model is stable and high-volume enough, and the near-monthly model release cadence is what holds it back.

Read More: contrast this with a viral chip "breakthrough" that turned out to be a marketing number.

Sources:

- AMD Acquires Taalas to Advance Compute Solutions for Rapidly Growing AI Inference Market (AMD Newsroom)
- AMD buys Taalas, startup that hardwires AI models into its silicon (CNBC)
- AMD acquires AI chip startup Taalas to boost inference performance by etching models into silicon (The Register)
- Taalas Etches AI Models Onto Transistors To Rocket Boost Inference (The Next Platform)

Disclaimer: *For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation.* ***Read more:*** */terms-of-service*