# 2025: Fundraising & Strategic Milestones — U.S. AI Inference Providers

This report covers key 2025 capital events and strategic agreements for leading U.S. AI inference / accelerator firms: Groq, Cerebras, SambaNova, Positron, and Etched. The focus is on publicly confirmed deals, valuations, and business implications.

## Groq

February 10, 2025 — Groq and the Kingdom of Saudi Arabia announce a $1.5 billion commitment for deployment of Groq’s LPU-based inference infrastructure in Saudi Arabia (not equity financing).

• This is a revenue / deployment contract rather than a fundraising round.

• Groq has indicated expectation of generating ~$500 million in 2025 from this commitment.

• The deal aligns with Saudi Arabia’s Vision 2030 push to build AI infrastructure.

September 2025 — Groq secures a $750 million equity financing, pushing its post-money valuation to $6.9 billion.

• The raise roughly more than doubles its prior valuation.

• Investors included Disruptive, BlackRock, Neuberger Berman, Deutsche Telekom Capital Partners, Samsung, Cisco, Altimeter, 1789 Capital, and Infinitum.

Analyst Take: Groq’s 2025 is characterized by a dual track: (1) a major infrastructure contract in Saudi Arabia that positions it as a global deployer of inference hardware, and (2) a sizable equity raise that underscores investor confidence in its growth and valuation trajectory. The Saudi deal provides a near-term anchor to support scale, while the funding underpins expansion of infrastructure and R&D.

## Cerebras

September 30, 2025 — Cerebras announces an oversubscribed Series G round of $1.1 billion, valuing the company at $8.1 billion post-money.

• The round was led by Fidelity Management & Research and Atreides Management, with participation from Tiger Global, Valor Equity Partners, 1789 Capital, Altimeter, Alpha Wave, and Benchmark.

• Funds are earmarked to expand AI processor / system research, U.S. manufacturing, and data center capacity.

April 2025 — Cerebras, in partnership with Canadian chip firm Ranovus, secures a $45 million DARPA / U.S. military contract to accelerate chip interconnect capabilities.

October 2025 (after Series G) — The company files to withdraw its planned U.S. IPO, citing its recent large financing round as a key motive for postponement.

Analyst Take: Cerebras’ 2025 fundraising and contract landscape reflect a maturing business model: deep capital infusion to scale operations and infrastructure, paired with government / defense engagements to anchor longer-term revenue. Its decision to pull back from an IPO post-Series G suggests a preference to expand privately until further scale is achieved.

## SambaNova

2025 — No new publicly disclosed equity rounds.

The company continues product development and deployment of inference infrastructure and API / cloud offerings, competing in the “token speed / latency / cost-efficiency” arena.

Analyst Take: SambaNova appears to emphasize execution, customer acquisition, and platform deployment rather than fresh fundraising in 2025. Its stability without new rounds may reflect sufficient capital buffer or a strategic shift to monetization.

## Positron

February 2025 — Positron raises $23.5 million (seed) to scale U.S.-manufactured inference accelerator production.

July 2025 — Positron completes a $51.6 million Series A (oversubscribed), intended for field deployment of its first-generation accelerator (Atlas) and R&D toward its next generation (Asimov).

Analyst Take: Positron is highly active in 2025, advancing from seed to growth funding within the year. Its capital plan is geared toward executing a hardware roadmap and entering production / deployment phases rapidly, positioning itself as a rising competitor in the inference hardware landscape.

## Etched

April 2025 — Etched secured $85 million in an A1 series round, valuing the company at $1.5 billion, a significant jump from its series A valuation.

The firm persists in developing its specialized transformer-focused ASIC (codenamed “Sohu”) and refining architecture / system stacks for inference workloads.

Analyst Take: Etched remains in a stealth / development-heavy stage. While smaller in scale than Groq, Cerebras, and SambaNova, the firm’s narrow architectural focus could pay off if execution and yield scale accordingly.

## Comparative Insights & Strategic Themes (2025)

1. Capital Leaders:

- Cerebras’ $1.1B Series G and Groq’s $750M raise are the largest disclosed equity infusions in 2025.

- Positron’s paired seed + Series A indicates strong investor conviction in new entrants.

2. Revenue / Deployment Deals:

- Groq’s $1.5B Saudi commitment stands out as a non-dilutive infrastructure agreement, anchoring its deployment confidence in a regionally strategic market.

- Cerebras’ DARPA contract offers balance between commercial and government-tied revenue streams.

3. Valuation Momentum & Investor Confidence:

- Groq’s valuation doubling (to $6.9B) reflects aggressive growth expectations.

- Cerebras’ $8.1B post-money valuation cements its status among the top-valued inference infrastructure players.

4. IPO / Exit Postures:

- Cerebras postponing its IPO after a major private raise suggests a deliberate approach to scaling privately first.

- Groq has not yet publicly filed for listing, likely continuing private growth for now.

5. Execution over Fundraising for Some:

- SambaNova’s quieter capital presence suggests a prioritization of deployment, product maturation, or conservatism in valuation timing.

6. Inference as the Differentiator:

- The narrative across all players emphasizes inference (token throughput, energy / latency efficiency, cost-per-inference) over general-purpose compute or training.

- The arms race is in throughput, power efficiency, latency, and deployability—not raw FLOPS alone.