AEVION QVenture — Investment Memo

Generated 2026-11-26 · AEVION AI Investment Analyst · not investment advice

FICTIONAL EXAMPLE: the company in this example is made up; any resemblance of names is coincidental. All figures are model output on invented inputs.

Nova Compute

AI Infrastructure / Tooling · seed · US · raising $6,000,000

Score 60.6/100 — WATCH (conviction: low)

Signal coverage: 44% company-specific (1 metrics parsed from the plan); remainder from sector priors.

Investment memo

Verdict: watch, not lead—engage with a small conditional check rather than anchoring the round. Nova has genuine early product-market signal (22 paying teams, 41% measured savings, 99.95% uptime over 90 days), and the strongest reason to lean in is that real, verifiable customer value at seed with clean execution is rare. But the single strongest reason against is structural and consistent across every lens and the quant model's 37/100 competitive headroom: Nova is a thin arbitrage layer squeezed between commoditizing model vendors and hyperscalers who can absorb cross-cloud routing natively, collapsing the very spread it monetizes. Undisclosed LTV/CAC, churn, and net-of-egress margins make the thesis unverifiable today. Entry plan: offer $1.3M for ~6% (hard-cap $2.1M), staged—release half now, half on three diligence gates: audited net-of-egress savings, NRR above 110%, and confirmation that spot-arbitrage does not breach cloud ToS. Reserve ~$2M for pro-rata. Size at ~1% of the fund; pass if margins or ToS fail.

Narrative engine: model anthropic (per-part source not recorded)

Entry strategy

Ticket: $1,317,240 target (range $658,620–$2,107,584)
Target ownership: 6%
Valuation band (pre-money): $8,254,400 / $15,954,000 / $31,908,000
Modelled return (not a forecast): 6.51x multiple (16x base) · horizon 7yr · loss prob 60%

Deployment schedule:

- 40% — Entry: On close, after founder + IP + cap-table diligence.
- 35% — Milestone: Product-market fit signal (retention cohort / first repeatable revenue).
- 25% — Pro-rata: Reserve for next priced round to defend ownership.

Portfolio: Size at ~1.1% of a diversified venture portfolio (fractional-Kelly, conviction-scaled). Reserve 1,975,860 USD for pro-rata follow-on.

Score breakdown

Market size & growth — 62/100 (weight 14%)
~$190B TAM, 19% CAGR (AI Infrastructure / Tooling).

Timing / tailwinds — 68/100 (weight 5%)
Sector growth 19% vs. 12% neutral baseline.

Moat / defensibility — 61/100 (weight 16%)
data scale is the category's mature moat (ceiling 78), but ~61% realized at seed given disclosed traction — an unproven moat is discounted toward the 35 "no demonstrated defensibility" floor.

Unit economics potential — 55/100 (weight 15%)
~65% mature gross margin, capital intensity 70% (sector reference).

Team / execution signal — 64/100 (weight 28%)
Quantified traction: $576k MRR.

Scientific / tech feasibility — 60/100 (weight 7%)
inference cost curves, sparse/MoE architectures, eval + safety tooling

Regulatory / legal headroom — 77/100 (weight 7%)
Regulatory intensity 35% (higher = more legal drag).

Competitive headroom — 37/100 (weight 8%)
Competitive intensity 90%. foundation-model commoditization compressing the middle layer.

Analyst council

Research Scientist — Solid systems engineering, not scientific novelty — 60/100 feasibility is fair; the real risk is durability, not buildability

+ The 60/100 feasibility score is well-calibrated, arguably slightly generous. This is applied distributed-systems engineering (spot-capacity arbitrage, SLA-aware scheduling, cross-cloud failover), not a research bet — there is no unsolved scientific problem here. The 99.95% measured uptime over 90 days and 41% average cost reduction across 22 teams demonstrate the core loop already works in product

+ The technical frontier they lean on (inference cost curves, MoE/sparse serving) actually cuts against them, not for them: per-token inference cost has been falling ~4-10x/year (cf. GPT-4-class pricing declines 2023-2025), and each foundation-model price cut shrinks the absolute dollars their 41% arbitrage captures. Their savings are a percentage of a rapidly deflating base — a structural headwind

+ Latency-SLA routing across five clouds' spot pools is genuinely hard at the tail: spot preemption (AWS gives 2-min warning, GCP 30s), cold-start on evicted GPU pools, and cross-region egress fees can silently erode the advertised 41% net-of-egress. The claimed uptime suggests they handle this, but 90 days / 22 teams is thin evidence for the p99.9 behavior enterprises underwrite.

+ Defensibility is the crux the feasibility score can't rescue: the routing logic is replicable, and the disclosed data-scale moat (61/100) requires request-volume network effects they haven't yet demonstrated at $48k MRR. Note the traction/plan mismatch — description says $48k MRR (~$576k ARR) while the model reads $576k as MRR; the lower figure is the honest read and reframes this as very early.

! Commoditization from above and below: cloud providers (Azure AI, Bedrock) and inference platforms (Together, Fireworks, OpenRouter) are shipping native multi-model routing; a cross-cloud router is a feature the hyperscalers can absorb, consistent with the 37/100 competitive-headroom score at 90% intensity.

! Deflating value pool: if per-token prices keep falling 4x+/year, a 41% relative saving on a shrinking bill yields declining absolute customer value and pricing power — the arbitrage spread that funds their take rate compresses toward zero.

! Unverified net economics: no LTV/CAC, no bottom-up TAM, and cost-reduction claims are gross of cross-cloud egress and their own margin. The stress test and TAM triangulation could not run — until net-of-egress savings and retention are disclosed, the 41% headline number is unaudited.

Data Analyst — 55/100 on unit economics is generous for a cost-arbitrage router with no LTV/CAC and margin-squeeze exposure

+ Push back on 55/100 as too kind, not too harsh: the ~65% mature GM is a sector reference, not observed. A router that captures a slice of a 41% cost saving faces structural GM pressure — if Nova prices as % of savings, its margin depends on a spread (spot GPU price vs SLA-met floor) that compresses as clouds add their own spot-optimization. Real GM could sit 40-55%, not 65%.

+ Traction is real but thin: $48k MRR / 22 teams = ~$2.2k ARPA/mo (~$26k ARPA/yr). No CAC, no churn, no gross retention disclosed — so LTV/CAC and payback are unknowable. STRESS TEST was not run for exactly this reason. At seed, 22 logos over 90 days is a signal, not a business.

+ Note the model's $576k 'MRR' is an annualized artifact ($48k x 12); do not double-count. On $576k ARR, a $6M raise implies ~10x forward ARR entry — reasonable for infra IF net revenue retention >120% and GM proves durable, expensive if churn is high or margin is thin.

+ TAM triangulation not run — the $190B TAM is the whole AI-infra market, not Nova's serviceable wedge. Bottom-up: inference spend routed through third-party optimizers is a fraction; SAM is more plausibly single-digit $B. 41% cost reduction is the strongest datapoint — it's a quantified, defensible value prop if it holds across workloads.

! Disintermediation / commoditization (competitive headroom 37, intensity 90): the five clouds Nova arbitrages can each ship native cross-pool spot routing, collapsing the arbitrage spread and Nova's GM to near zero. This is the thesis-killer and it's structurally disclosed.

! Unit economics are undisclosed — no CAC, churn, NRR, or contribution margin. Without these the 55 score is a placeholder. Kill signal: gross churn >20% or blended GM <45% at scale.

! Value-capture fragility: if Nova prices on savings, customers renegotiate as savings shrink; if flat-fee, it competes with free/native tooling. Either path caps ARPA expansion.

Economist — Real cost savings, but a thin-margin arbitrage layer squeezed between commoditizing model vendors and clouds absorbing the routing feature

+ The 62/100 market score is defensible but flatters this business. The ~$190B AI-infra TAM at 19% CAGR is real, yet Nova's serviceable slice is narrow: it monetizes the delta between GPU pools, not compute itself. Take-rate on a 41% cost cut across five clouds' spot markets is a fraction of a fraction — I'd model SAM closer to single-digit billions, so treat 62 as a TAM ceiling the company can't ca

+ Demand is genuinely elastic here — inference spend is a top-3 cost line for LLM-heavy teams, so a verified 41% reduction sells itself, evidenced by 22 paying teams and 99.95% uptime. That's the bull case: the product works and the pain is quantified. Elastic demand cuts both ways though — the same buyers will switch the instant a cheaper router (or a free cloud-native one) appears.

+ Moat at 61 is too generous for what's disclosed. The routing logic is replicable; the only durable asset is placement-optimization data at scale, and 22 teams / $48k MRR is nowhere near the flywheel. Multi-cloud spot integration is a 6-9 month engineering lift, not a structural barrier — this reads closer to the 35 'undemonstrated defensibility' floor than to 61.

+ Note the traction discrepancy: brief states $48k MRR (~$576k ARR) while the model's team line reads '$576k MRR' — a 12x conflation. At $48k MRR the business is pre-scale; the 64 team score built partly on a misread ARR figure should be discounted.

! STRUCTURAL — margin compression from both ends: the disclosed 65% gross margin assumes clouds keep discrete spot pricing gaps. AWS/GCP/Azure have every incentive to ship native cross-region/spot routing (or foundation vendors bundle inference optimization free), collapsing the arbitrage Nova monetizes. Competitive headroom at 37 (90% intensity) is the single most honest number in this model.

! Capital intensity flagged at 70% plus undisclosed LTV/CAC (stress test not run) is a real hole. Selling a switching-cost-light cost-savings tool implies high churn risk and expensive land-and-expand; without CAC payback I cannot verify the unit economics survive contact with enterprise sales cycles.

! Key-man dependency on continued multi-cloud API access — if a cloud rate-limits or contractually restricts spot arbitrage resellers, Nova's five-cloud failover premise degrades to a subset, eroding both the 41% savings claim and the failover value prop.

Corporate & Regulatory Lawyer — 77/100 legal headroom is fair but flatters a middle-layer router carrying multi-cloud contract, export-control, and data-flow exposure

+ The 77/100 is directionally right — Nova is sectoral-only (FTC §5, no AI statute, no financial-services license), so no gating regulator can block operations. But 77 undersells contractual and export risk that lives outside 'regulatory intensity': the real legal surface here is B2B and cross-border, not agency licensing.

+ Core structural exposure is the five-cloud dependency. Nova routes across spot capacity from AWS/GCP/Azure/etc — most hyperscaler ToS restrict resale/brokerage of compute and prohibit circumventing pricing. A single provider revoking API access or invoking anti-resale clauses breaks the '41% cost reduction' thesis overnight. Diligence must pull every underlying cloud agreement and confirm Nova is

+ Data/privacy is live because Nova sits in the inference request path: prompts/completions transit its router. Absent a federal omnibus law, CCPA/CPRA plus HIPAA/GLBA apply if any of the 22 teams push regulated data. Nova needs signed DPAs, sub-processor disclosures, and a no-retention/no-training posture in writing, or it inherits liability as a processor. Export controls (EAR advanced-compute rul

+ Round structure is clean and standard: 506(b) or 506(c) Reg D, Form D within 15 days, 1x non-participating pref, broad-based weighted-average anti-dilution, pro-rata + information rights, board observer. At $6M seed a post-money SAFE or priced round is market; insist on IP assignment reps and confirmation all founder/contractor code is assigned to the company.

! Anti-resale / ToS violation: if Nova's spot-arbitrage across five clouds breaches hyperscaler terms, providers can terminate access and the product's cost-savings moat and 99.95% uptime claim collapse — this is a going-concern legal risk masked by the benign 77 score.

! Uncapped downstream liability: an SLA-driven router promising latency guarantees and failover invites breach-of-SLA and consequential-damages claims from paying teams if a misroute causes an outage; without liability caps and clear data-processing terms Nova carries disproportionate exposure for a seed-stage balance sheet.

! IP thinness: the routing logic is likely trade-secret/orchestration know-how, not patented — commoditization risk (quant flags this, competitive headroom 37) means defensibility is contractual and operational, not legal; confirm no lapsed filings and that key algorithms are protected as trade secrets with enforceable employee NDAs.

Market data sources

- MarketsandMarkets (2026) — AI infrastructure $135.8B (2024) !' $394.5B by 2030 at 19.4% CAGR (~$190B in 2026)
  https://www.marketsandmarkets.com/Market-Reports/ai-infrastructure-market-38254348.html
- Precedence Research (2026) — To reach $465.9B by 2034
  https://www.precedenceresearch.com/artificial-intelligence-infrastructure-market

Assumptions & limitations

- Market size / growth for AI Infrastructure / Tooling is anchored to MarketsandMarkets (2026): AI infrastructure $135.8B (2024) !' $394.5B by 2030 at 19.4% CAGR (~$190B in 2026). Full citations are listed under "Market data sources".
- Signal coverage: ~44% of the score is backed by the plan's own disclosed metrics (1 quantified field); the remainder uses AI Infrastructure / Tooling sector priors — add financials to raise it.
- Stage norms reflect US-market seed deals; adjust for geography "US".
- Score is a screening signal, not a substitute for legal, financial, and technical due diligence.

This memo is generated by an AI screening tool for research purposes and is not investment advice, an offer, or a solicitation. Figures are model estimates, not guarantees.