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.

TutorPath

AI Applications (vertical SaaS) · pre-seed · US · raising $1,500,000

Score 63.4/100 — WATCH (conviction: medium)

Signal coverage: 0% company-specific (0 metrics parsed from the plan); remainder from sector priors.

Investment memo

Verdict: watch-and-conditionally-lead — a small anchor check now, gated on diligence, rather than a pass or a full commit. The strongest reason for is a rare pre-seed efficacy signal (+0.8 grade-level gain across 4,200 paid students in real districts), pairing genuine demand-side pull with a credible misconception-diagnosis wedge. The strongest reason against is that the entire thesis rests on one uncontrolled pilot number with 0% disclosed metrics: if the diagnosis is prompt-engineering over a foundation model rather than a proprietary error-taxonomy, an incumbent (Khan/IXL) or the model vendor collapses the 41/100 competitive headroom to zero. Entry plan: lead $640K for ~8% at a ~$6.5M pre-money, hard-cap total exposure at $750K, and reserve ~$950K for pro-rata. Stage the money against two diligence gates before wiring — (1) diagnostic-accuracy data plus a control-group/RCT commitment to clear ESSA tiers, and (2) FERPA/COPPA DPAs and disclosed LTV/CAC. No proof on either, no deal.

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

Entry strategy

Ticket: $638,592 target (range $319,296–$750,000)
Target ownership: 8%
Valuation band (pre-money): $3,120,600 / $6,482,400 / $12,964,800
Modelled return (not a forecast): 7.45x multiple (23.8x base) · horizon 8yr · loss prob 69%

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 ~2.3% of a diversified venture portfolio (fractional-Kelly, conviction-scaled). Reserve 957,888 USD for pro-rata follow-on.

Score breakdown

Market size & growth — 57/100 (weight 14%)
~$70B TAM, 36% CAGR (AI Applications (vertical SaaS)).

Timing / tailwinds — 100/100 (weight 5%)
Sector growth 36% vs. 12% neutral baseline.

Moat / defensibility — 53/100 (weight 16%)
switching costs is the category's mature moat (ceiling 74), but ~45% realized at pre-seed given disclosed traction — an unproven moat is discounted toward the 35 "no demonstrated defensibility" floor.

Unit economics potential — 69/100 (weight 15%)
~70% mature gross margin, capital intensity 35% (sector reference).

Team / execution signal — 58/100 (weight 28%)
commercial validation cited

Scientific / tech feasibility — 98/100 (weight 7%)
agentic workflows, domain eval harnesses, retrieval + tool orchestration

Regulatory / legal headroom — 74/100 (weight 7%)
Regulatory intensity 40% (higher = more legal drag).

Competitive headroom — 41/100 (weight 8%)
Competitive intensity 85%. thin wrapper risk — value must accrue above the model layer.

Analyst council

Research Scientist — Misconception-diagnosis is scientifically credible but 98/100 overstates the frontier difficulty here — the hard part is measurement, not the model

+ I push back on 98/100. The engineering stack (retrieval + tool orchestration + eval harness) is real but not novel or defensible — it is a standard 2024-era agentic pattern any competent team can replicate in weeks. A near-perfect feasibility score conflates 'technically buildable' with 'technically differentiated'; the science here is easy, which is precisely why the competitive headroom is 41/10

+ The genuinely credible science is misconception diagnosis: HS math has a well-catalogued taxonomy of ~50-200 recurring error patterns (e.g., Ashlock's error-pattern research, VanLehn's 'buggy' subtraction rules, Newman error analysis). Mapping wrong answers to a specific bug is a tractable, decades-studied problem — this is the one area where a proprietary labeled dataset could accrue value above

+ The +0.8 grade-level gain over one semester is a strong headline but scientifically unverified: no control group, n and confounders undisclosed, and 'grade-level gain' measurement instrument unstated. Bloom's 2-sigma tutoring literature and ITS meta-analyses (e.g., VanLehn 2011, effect sizes ~0.4-0.75) make +0.8 plausible but not extraordinary — it needs an RCT-style comparison to survive district

+ De-risking breakthrough: a validated, model-agnostic misconception-classification layer with published accuracy (e.g., >85% top-1 bug attribution vs. expert-labeled ground truth) plus an ESSA Tier II/III quasi-experimental study. That converts a thin wrapper into a defensible diagnostic asset and satisfies district procurement evidence bars.

! Thin-wrapper realized: if diagnosis is prompt-engineering over a frontier LLM rather than a proprietary error-taxonomy model, a foundation-model vendor or Khan/IXL incumbent replicates it, collapsing the 41/100 competitive headroom to zero.

! Efficacy claim fails replication: +0.8 gain may reflect selection/novelty/teacher effects; absent a controlled study it will not clear ESSA evidence tiers, stalling district sales that increasingly require Tier I-III proof.

! No disclosed unit economics or diagnostic accuracy metrics (0% signal coverage) means the core scientific claim — that it identifies the *specific* misconception — is entirely unaudited; the whole thesis rests on one unverified pilot number.

Data Analyst — 69/100 unit-econ score is a sector-prior placeholder — zero disclosed metrics; district SaaS margins support it, but sales cycle can wreck payback

+ The 69/100 rests entirely on sector priors (~70% GM, 35% capex) — the plan disclosed 0 quantified financial fields, so this score is unearned, not observed. Software GM of 70-80% is plausible for edtech, but I'd hold judgment: no ACV, no CAC, no LTV, no churn is disclosed. Real district-SaaS gross margin often runs 55-65% once you load implementation, PD/training, and customer-success headcount th

+ CAC/payback is the actual risk district SaaS hides. K-12 sales cycles run 6-18 months with committee/procurement/budget-cycle gates; blended CAC of $15-40k per district is typical. At 3 pilots the company hasn't proven a repeatable motion — I'd want ACV per district (likely $10-50k), gross retention, and pilot-to-paid conversion before crediting LTV. Multi-year contracts help LTV, but budget churn

+ TAM/SAM/SOM is unfalsifiable as presented: $70B is the whole vertical-SaaS sector, not HS-math intervention. Bottom-up: ~13k US districts x ~$30-80k intervention budget for math implies a SAM closer to $0.5-2B — an order of magnitude below the headline, and enough to build a real business but not the $70B narrative.

+ Traction is the strongest datapoint: +0.8 grade-level gain in the largest pilot is a credible efficacy claim IF measured against a control and validated (ESSA evidence tier matters for district buying). 4,200 active students across 3 pilots is real usage. This is what a pre-seed round should be funding — converting efficacy into a repeatable, marginable sales motion.

! Thin-wrapper / margin compression: value must accrue above the model layer (comp intensity 85%). If the misconception-diagnosis logic is a prompt over a foundation model, per-student inference cost + a free/cheap incumbent (Khanmigo, IXL, publisher tools) caps both pricing power and the assumed 70% GM.

! Payback undisclosed and likely brutal: with 6-18mo K-12 cycles and ESSER funding expiring, if payback exceeds 18-24 months on a pre-seed balance sheet, $1.5M won't reach a repeatable motion. No LTV/CAC = no stress test possible; this is the single metric that confirms or kills the thesis.

! Efficacy claim unverified: +0.8 grade-levels may lack a control group or independent validation; district procurement increasingly requires ESSA Tier 1-2 evidence. If the gain doesn't survive an RCT, the core sales wedge collapses.

Economist — Real learning-gains signal, but $70B TAM is the wrong denominator — district-intervention budget is a fraction of it

+ The 57/100 market score is generously anchored to a $70B vertical-SaaS TAM; the serviceable market is narrower. US high-school math is ~15M students; district intervention/supplemental spend runs ~$30-80/student/year, implying a bottom-up SAM closer to $0.5-1.2B, not $70B. I'd push the score DOWN on size but the 36% CAGR tailwind is real — net wash around mid-50s.

+ Demand is inelastic-ish but budget-gated: districts buy on ESSER/Title I dollars and cycle procurement annually. ESSER funds expired Sept 2024, so the tailwind that inflated ed-tech demand 2021-24 is now reversing — this is a macro headwind the timing=100 score ignores entirely.

+ The +0.8 grade-level gain is the single strongest asset — if it survives an RCT, it converts to defensible efficacy claims that justify premium pricing and create switching costs (score's 53 moat is fair, ~45% realized). Efficacy evidence, not the model, is where economic rent accrues here.

+ Thin-wrapper risk is acute: 'diagnose misconception from wrong answers' is increasingly a base-model capability.

The moat must be the proprietary misconception taxonomy + labeled error-response dataset from 4,200 students, plus district distribution — not the LLM orchestration, which the 98 feasibility score correctly flags as commoditized.
! ESSER cliff: the primary district funding source for interventions largely lapsed in 2024, compressing the budget pool precisely as TutorPath scales. Sales cycles of 6-12 months against shrinking budgets can starve a pre-seed on $1.5M.
! Efficacy is n=1 (largest pilot, no control group, one semester). +0.8 grade-levels may reflect selection or novelty effects; without an RCT the core value prop is unproven and competitors (Khan/Khanmigo, IXL, incumbent LMS vendors) can claim parity for free or bundled.
! Value-accrual above the model layer is undemonstrated: if OpenAI/Anthropic ship better native math tutoring, the wrapper's gross margin (~70%) and pricing power erode. Competitive headroom of 41 understates the free/bundled-incumbent threat in K-12.

Corporate & Regulatory Lawyer — 74/100 legal headroom is too generous — child-data (FERPA/COPPA) exposure and AI efficacy claims are the real drag at 40% intensity

+ Push back on 74/100: the score treats this as generic vertical SaaS at 40% regulatory intensity, but selling to K-12 districts puts TutorPath squarely under FERPA (34 CFR Part 99) and COPPA (children <13) — the district acts as the FERPA gatekeeper and TutorPath is a 'school official' vendor bound by direct-control and data-use limits. Undisclosed whether DPAs, the SDPC national data-privacy agree
+ Efficacy-claim liability is the sharpest FTC §5 exposure: '+0.8 grade-level gain' is an advertised outcome claim from an n=1 largest-pilot with no control group. FTC AI-claims guidance and the 'Operation AI Comply' posture mean unsubstantiated learning-gain marketing is a deception risk — this claim must be caveated (single cohort, no counterfactual) in sales collateral or it becomes a warranty in
+ Deal structure: at $1.5M pre-seed, use a post-money SAFE or priced Reg D 506(b) round (avoid 506(c) verification friction with a small accredited syndicate). Insist on pro-rata rights, information rights, and a board observer seat. Given the thin-wrapper structural risk, add IP reps that the misconception-diagnosis logic and eval harness are owned/assignable — not merely prompt-engineering over a
+ IP posture is honestly thin at pre-seed: the defensible asset is the labeled misconception taxonomy and student-response dataset (4,200 students), not the model. That data is also the liability — FERPA restricts using student PII to train models absent district authorization. Diligence must confirm training-data rights are contractually secured; otherwise the moat and the compliance risk are the s
! FERPA/COPPA non-compliance: if TutorPath trained or improved models on identifiable student data without proper district DPAs, districts can terminate and the FTC/state AGs (COPPA carries penalties up to ~$53k per violation) can act — killing the sales channel and the data moat simultaneously.
! Deceptive-claims exposure: the +0.8 grade-level figure, if used in marketing without disclosing it is single-pilot and uncontrolled, is a live FTC §5 and state-UDAP risk, and could be recast as a breach-of-warranty claim by a district that underperforms.
! Emerging state AI patchwork: Colorado AI Act (effective 2026) and similar bills may classify education AI affecting student outcomes as 'high-risk,' triggering impact-assessment and disclosure duties that raise compliance cost as the company scales across states.

Market data sources

- Global Market Insights (2026) — Generative AI $83.3B in 2026 !’ $988.4B by 2035 at 31.6% CAGR
  https://www.gminsights.com/industry-analysis/generative-ai-market
- Fortune Business Insights (2026) — Generative AI $161B in 2026 !’ $1,260.2B by 2034 at 39.6% CAGR
  https://www.fortunebusinessinsights.com/generative-ai-market-107837

Assumptions & limitations

- Market size / growth for AI Applications (vertical SaaS) is anchored to Global Market Insights (2026): Generative AI $83.3B in 2026 !’ $988.4B by 2035 at 31.6% CAGR. Full citations are listed under "Market data sources".
- Signal coverage: ~0% of the score is backed by the plan's own disclosed metrics (0 quantified fields); the remainder uses AI Applications (vertical SaaS) sector priors — add financials to raise it.
- Stage norms reflect US-market pre-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.