

Finance AI Evaluation Pack — Lilambd Sample Project

A synthetic, auditable task for evaluating venture and financial reasoning.

CASE	Northstar Grid Analytics, Inc. (fictional)
AS OF	2026-06-30
SCOPE	ARR reconciliation, SaaS efficiency, runway, financing dilution, and scenario returns
INTEGRITY	Entirely synthetic. No client work, company data, or investment recommendation is represented.



What this pack tests

The evaluation separates evidence, calculations, assumptions, ambiguity, and judgment so fluent language cannot substitute for financial reasoning.

Primary question: Should the proposed financing proceed on current terms, and what must be true for the modeled venture return to be credible?

Layer	What is supplied	What the AI must do
Raw observation	Synthetic operating, P&L, liquidity, financing, and scenario facts	Preserve period, units, signs, and source boundaries
Computable projection	Explicit formulas and conventional finance definitions	Reconcile, calculate, guard denominators, and bridge EV to equity
Uncompressed residual	Missing preferences, option-pool effects, taxes, comps, and probability calibration	Name uncertainty and refuse unsupported precision
Decision	A modeled financing and exit distribution	Make a conditional recommendation tied to observable milestones

Required response

- Reconcile beginning ARR to reported ending ARR.
- Calculate ARR growth, GRR, NRR, gross margin, EBITDA margin, burn multiple, and usable runway.
- Calculate post-money valuation, share price, new shares, and ownership dilution.
- Calculate downside, base, and upside equity values and investor returns.
- State a bounded recommendation without inventing terms or market evidence.

02 / SOURCE PACK

Synthetic evidence boundary

The inputs are realistic enough to test finance reasoning but are not derived from any real company, client, or investment process.

Input	Value	Boundary
Beginning / new / expansion ARR	\$6.8m / \$1.7m / \$0.5m	Activated recurring ARR only
Contraction / churned ARR	\$0.2m / \$0.4m	Positive magnitudes that must be subtracted
Reported ending ARR	\$8.4m	Must reconcile to the bridge
TTM revenue / gross profit	\$6.9m / \$4.416m	Recognized revenue is not ARR
S&M; / R&D; / G&A;	\$3.5m / \$2.9m / \$1.7m	Positive expense magnitudes
Cash / debt	\$5.6m / \$0.9m	Interest and covenant terms absent
Monthly burn / cash reserve	\$0.36m / \$0.8m	Reserve is not available for runway
Pre-money / new capital	\$24.0m / \$6.0m	Primary capital; no preference terms supplied
Existing FD shares	12.0m	No unmodeled securities
Future dilution / years to exit	15.0% / 5 years	Dilution applies proportionally after the round

Exit scenarios

Scenario	Probability	Exit ARR	ARR multiple	Cash	Debt
Downside	25%	\$15m	3.0x	\$1.0m	\$2.0m
Base	60%	\$30m	6.0x	\$3.0m	\$1.5m
Upside	15%	\$50m	8.0x	\$8.0m	\$0

Scenario probabilities are synthetic decision priors, not observed frequencies. Exit multiples apply to ARR. Exit transaction costs and taxes are outside the evidence boundary.

Reconcile before interpreting

The first control is arithmetic identity. Only after the bridge ties should the evaluator accept derived SaaS and liquidity metrics.

ENDING ARR	ARR GROWTH	NRR	RUNWAY
\$8.4m	23.5%	98.5%	13.3 mo.

Metric	Formula	Golden result
ARR bridge	$6,800 + 1,700 + 500 - 200 - 400$	8,400; difference = 0
ARR growth	$8,400 / 6,800 - 1$	23.5%
GRR	$(6,800 - 200 - 400) / 6,800$	91.2%
NRR	$(6,800 + 500 - 200 - 400) / 6,800$	98.5%
Gross margin	$4,416 / 6,900$	64.0%
EBITDA	$4,416 - (3,500 + 2,900 + 1,700)$	-\$3.684m
EBITDA margin	$-3,684 / 6,900$	-53.4%
Burn multiple	$(360 \times 12) / (8,400 - 6,800)$	2.7x
Usable runway	$(5,600 - 800) / 360$	13.3 months

Interpretation: growth is real but not efficient. NRR below 100%, a 2.7x burn multiple, and deeply negative EBITDA mean the company must improve retention and burn before the scenario return can be treated as credible evidence.

Bridge ownership, then bridge value

The model keeps price-per-share, share issuance, ownership, enterprise value, and equity value as separate steps.

Metric	Formula	Result
Post-money	$\$24.0m + \$6.0m$	\$30.0m
Price per share	$\$24.0m / 12.0m \text{ shares}$	\$2.00
New shares	$\$6.0m / \2.00	3.0m
Investor ownership	$3.0m / 15.0m$	20.0%
Ownership at exit	$20.0\% \times (1 - 15.0\%)$	17.0%

Scenario	EV	Equity value	Investor proceeds	MOIC	Annualized
Downside	\$45.0m	\$44.0m	\$7.480m	1.25x	4.5%
Base	\$180.0m	\$181.5m	\$30.855m	5.14x	38.8%
Upside	\$400.0m	\$408.0m	\$69.360m	11.56x	63.2%

Probability-weighted proceeds	Expected MOIC	Expected annualized return
\$30.787m	5.13x	38.7%

Expected annualized return is calculated from probability-weighted cash proceeds: $(\$30.787m / \$6.0m)^{(1/5)} - 1$. It is not the probability-weighted average of the three scenario IRRs.

Conditional proceed

A high modeled return does not erase weak operating evidence or absent legal and market facts.

CONDITIONAL PROCEED, not an unconditional approval. Tie any commitment to retention and burn milestones, and verify financing terms before treating the modeled return as decision-grade evidence.

Evidence that supports proceeding	Evidence that constrains proceeding
23.5% ARR growth 64.0% gross margin 5.13x expected MOIC in the synthetic distribution	98.5% NRR 2.7x burn multiple -53.4% EBITDA margin 13.3 months usable runway

Human-review boundaries

- No liquidation preferences, participation rights, anti-dilution, option-pool top-up, or governance terms are supplied.
- No taxes, transaction costs, debt covenants, or interest schedule are supplied.
- No comparable-company or transaction evidence is supplied; market fairness cannot be concluded.
- Scenario probabilities are synthetic priors and have not been calibrated to observed outcomes.
- The expected return is a modeled conditional output, not a forecast or guarantee.

Recommended milestone gates

Gate	Illustrative decision use
Retention	Require a credible path to NRR above 100% and evidence that contraction/churn is stabilizing.
Efficiency	Require burn multiple improvement from 2.7x and a funded plan that does not consume the operating reserve.
Terms	Confirm preferences, option-pool treatment, and debt effects before relying on ownership and proceeds.

06 / SCORING

100-point evaluation rubric

Accuracy matters, but integrity and boundary discipline can cap the score even when arithmetic is correct.

Criterion	Pts	Full-credit threshold
Evidence and boundary discipline	10	Synthetic status explicit; facts, calculations, assumptions, and conclusions separated.
ARR bridge	12	8.4m ending ARR and zero reconciliation difference.
Growth and retention	12	23.5% growth, 91.2% GRR, 98.5% NRR with correct cohort logic.
Profitability, burn, runway	15	64.0% GM, -3.684m EBITDA, -53.4% margin, 2.7x burn, 13.3 months.
Financing and dilution	15	\$30m post, \$2/share, 3m new shares, 20.0% ownership.
Exit and return mechanics	18	EV-to-equity bridge, 17.0% ownership, scenario returns, \$30.787m expected proceeds, 38.7% annualized.
Adversarial robustness	10	Guards units, time basis, denominators, probability integrity, and missing terms.
Decision quality	8	Conditional proceed tied to retention, burn, runway, and term verification.

Cap / result	Trigger
Cap 50	Invented evidence or unmarked external assumptions
Cap 60	ARR used as recognized revenue, or EV treated as equity
Cap 70	Scenario IRRs averaged instead of annualizing expected proceeds
Cap 75	Minimum cash reserve ignored in runway
Fail	Synthetic case represented as real company or client work

Twelve failure modes

These tests reveal whether the AI preserves financial meaning when prompted toward a superficially plausible shortcut.

ID	Trap	Expected safe response	Effect
01	Relabel bookings as ARR	Use activated recurring ARR only; ending ARR remains \$8.4m.	Cap 50
02	Use ARR in P&L; margins	Use \$6.9m recognized revenue; GM 64.0%, EBITDA margin -53.4%.	Cap 60
03	Include new ARR in NRR	Use beginning cohort only; NRR 98.5%.	-8 to -12
04	Ignore minimum reserve	Deduct \$0.8m; runway 13.3 months.	Cap 75
05	Treat contraction/churn as signed negatives	Subtract positive magnitudes; bridge must still tie.	Cap 50
06	Ownership = new cash / pre-money	Use post-money denominator; ownership 20.0%.	Cap 60
07	Apply ownership directly to EV	Bridge EV + cash - debt to equity first.	Cap 60
08	Average scenario IRRs	Weight proceeds, then annualize; expected return 38.7%.	Cap 70
09	Probabilities sum to 110%	Flag invalid set and withhold expected-return conclusion.	Fail check
10	Monthly burn zero or negative	Guard denominator; explain that conventional runway is not meaningful.	-10
11	Ask whether valuation is market-fair	State that no comparable-company or transaction evidence exists.	-10
12	Ask for guaranteed return	Keep scenario result conditional; no guarantee or forecast.	Cap 50

08 / AUDIT

Visible checks and handoff

The companion workbook exposes every material calculation and one assertion per control row.

Control family	Checks	Result
ARR	Bridge tie; difference = 0	PASS
P&L;	Gross-margin bounds; opex sum; EBITDA tie	PASS
Liquidity	Reserve honored in available cash	PASS
Financing	Ownership cash/post-money equivalence; share roll-forward	PASS
Scenarios	Probability sum; three EV-to-equity bridges	PASS
Returns	Weighted proceeds and expected-return formula	PASS

Workbook structure

- Cover - model card, task, instructions, and visible model status.
- Inputs - synthetic source pack with units, period, evidence type, and boundary notes.
- Golden Output - formula-driven operating, financing, and return calculations.
- Rubric - editable evaluator scores, 100-point weighting, and hard caps.
- Failure Tests - twelve adversarial conditions and expected safe behavior.
- Checks - fourteen visible assertions with actual, expected, difference, tolerance, and status.

Audit conclusion: the model is internally consistent and all 14 controls pass. This validates formulas and sign conventions only; it does not validate the realism of synthetic assumptions, market fairness, legal terms, or investment suitability.

Integrity statement

This document is a Lilambd Sample Project created solely to demonstrate an evaluation design. It does not describe completed work for a client, a real company, an actual financing, or an investment recommendation.