AGI ALPHA FLYWHEEL ∞

When Paid Work Teaches the Network

Profitable delivery. Selective Alpha.
Expanding productive possibility.

An executed skill lifecycle, a matched commercial simulation,
and a numerical transformation witness

The result. Competent Beta is profitable under the modeled primary markets. Its quality requirements frequently fail. The selected retention policy keeps nothing. Cheaper retention produces useful peer reuse and some positive individual skill balances, but the complete retention program remains uneconomic.

TRANSFORMATIONAL INTELLIGENCE

Learning to change the game

Find your Alpha. Build your successor Ω.

Prepared for Vincent Boucher / MONTREAL.AI

Research execution and manuscript prepared with ChatGPT

27 October 2026 UTC | Evidence date: 26 October, Toronto

Research edition; no external peer-review or production qualification claimed. Model performance, identities, customers and payments are simulated. Software execution and ordinary-tool numerical results are identified separately.

Abstract

Can an automated enterprise earn useful operating profit from competent current tools, and then obtain additional value by retaining what paid work taught it? We implement a certified numerical-reformulation foundry around an existing, byte-pinned skill lifecycle. Candidate construction records, evidence gates, current-state authorization, graph transitions, resource constraints, finite-cash accounting and bounded policy selection execute as software; frontier-model outcomes, customers, transfer and payments are modeled. Development selects one common commercial controller and optional private/network retention policies before fresh confirmation. Across three 64-world primary cohorts, with 52 selling and 20 runoff weeks, Beta mean profit is $121,004, $919,696 and $1,810,306. Only 5, 44 and 39 worlds, respectively, meet the complete service-quality gate. All selected retention policies choose off, making primary incremental effects exactly zero. A prespecified enabled-assessor challenger loses money relative to Beta in each profile. A separate cheaper-retention mechanism cohort achieves 123 useful reuses, including 86 by peers, yet its mean direct retention balance is $-1,020 and its enterprise increment is $-139,215. Twenty-one individual admitted skills have positive directly linked balances before shared discovery costs. A separate executed numerical witness ties a retained template with competent structured Beta. All 5,082 histories reconcile, and nine complete records replay byte-for-byte. These results support profitable simplicity under stated assumptions and selective evidence requirements; they do not establish empirically profitable autonomy or a positive network flywheel.

Three questions, three answers. **Can Beta earn money?** Yes, conditionally in the primary simulation; quality qualification is incomplete. **Can paid work leave a useful reusable method?** Yes in a mechanism diagnostic, and the ordinary-tool witness exhibits a concrete representation. **Does the selected network policy add net enterprise value?** No positive increment is established. Retention off is selected; enabled alternatives incur losses.

1 Question, value and evidence

The commercial foundation is useful delivery. An enterprise does not need an accumulating skill graph to justify profitable use of existing frontier models and ordinary tools. Conversely, a profitable enterprise does not establish that its retained skills, network distribution or reinvestment create value. We separate these claims before choosing a venture, a price policy or a statistical analysis.

Beta is competent use of capabilities already available to the recipient: current frontier reasoning, ordinary numerical libraries, grounded decomposition, executable checking, critique, correction and accumulated ordinary task familiarity. **Alpha** is additional net mission advantage over that baseline and the recipient's existing inventory, within a declared scope. Familiar mathematics may support a useful transformation; familiarity also makes it available to Beta. A stored wrapper, graph edge, download or authorization does not by itself establish Alpha.

Let $\Pi_{w,p}$ be complete ordinary operating profit in world w under policy p. The primary commercial quantity is $E[\Pi_{w,\text{Beta}}]$, accompanied by a world-level quality indicator. The paired incremental quantities are

$$\Delta_{\text{private}} = E[\Pi_{w,\text{private}} - \Pi_{w,\text{Beta}}], \tag{1}$$

$$\Delta_{\text{share}} = E[\Pi_{w,\text{network}} - \Pi_{w,\text{private}}], \tag{2}$$

$$\Delta_{\text{total}} = E[\Pi_{w,\text{network}} - \Pi_{w,\text{Beta}}]. \tag{3}$$

These compare complete enterprises with matched market conditions, common commercial controls and the same initial resources. An individual capability's directly linked balance is a different quan-

tity. Its positive value can coexist with an unprofitable search program that paid for many rejected candidates.

The study executes a hybrid experiment. Table 1 identifies the evidence boundary. A software invariant is evidence about implementation, not an estimate of frontier performance or willingness to pay. World-level observations use hidden correctness for evaluation; operating policies cannot read that oracle.

Table 1: Evidence categories remain distinct throughout the package.

Category	What the study establishes
Executed software	Job/cost records; policy selection; A/B/C ledger; native admission, access and retirement; independent reconstruction; exact replay.
Executed numerical work	Fresh constrained problems, three conventional solver implementations, separate residual checks, stale-result rejection and composition fixtures.
Modeled outcomes	Astra/Terra task success, automated validator errors, candidates, customers, demand, transfer, obsolescence, collections, refunds and operating cash.
Modeled identity	Direct-label `*.agent.agi.eth` membership, ownership changes and stale authorization fixtures. No live ENS or production authority.
External source facts	Dated published model prices/configuration limits; a dated human-service rate schedule; source commits and supplied foundations. Task-matched performance remains unmeasured.

2 System reuse and critical prior evidence

We inspected the four requested MontrealAI repositories at fixed commits and coordinated through issue #4850 [14]. The research directory is additive. It changes no production registry, owners, active release paths, job settlement or `*.club.agi.eth` roles. The Jobs and Manager discovery interfaces explicitly describe synthetic local inspection without externally provisioned execution or settlement; their AGIALPHA18 and USDC6 routes must not be confused with this study's USD accounting. Fleet's current-state admission and refusal semantics inform the boundary; no offline fixture commissions a live worker. Exact commits, inspected paths and source hashes are in the evidence companion and `evidence/source_provenance.json`.

Eleven dependency files are reused byte-for-byte from the existing #4951 research lifecycle, commit `2727262caa52766a78916f2ed81f5afaf668edce`. The wrapper calls that native JavaScript implementation inside every simulated world. It does not merely serialize a diagram or replay a lifecycle after economic outcomes have already been decided. Original licenses and origins are preserved. Cryptographic hashing, typed commitments and local authenticated encryption execute; fixture attestations do not establish production identity or truth.

The supplied *Transformational Intelligence*, Successor Ω, PROOF LOOPS v2 and Merkle V114 foundations distinguish changed representation, executable realization and purpose-preserving outcome evidence from increased search or graph size [1, 2, 3, 4]. They also distinguish historical evidence from current authority. We apply those distinctions without claiming a new proof level, mainnet readiness or independent institutional certification.

Prior studies remain separate evidence populations (Table 2). The source review inspected their PR descriptions and selected committed outputs, not every historical raw ledger. Reported figures retain that status. We preserve conditional Beta successes, negative retention and service-quality failures. The positive paid-work report is a reason to test the mechanism, not permission to insert its unmeasured

time or quality benefits into a new model.

Table 2: Prior findings are retained under their own horizons and accounting. Amounts are approximate USD; no pooled headline is computed.

Study	Population / horizon	Reported commercial finding	Retention or quality qualification
#4950	64/profile; 104+16 weeks	Beta means 0.411m / 3.702m / 6.763m	Network increments approximately $-13,947/-12,063/-8,766$; favorable interval crosses zero.
#4951	128/profile; 104 selling + runoff	Beta means $-28,573/497,058/2,423,231$	Selected primary conducts no capability campaigns. Separate admitting trace loses at enterprise level.
#4952	Reference 14 worlds; 72 + runoff	Reference mean 151,024	Increment -64.74; 5/14 reference worlds exceed the stated false-pass bound.
Paid-work report	96/profile; 52+16 weeks	Beta means 85,636/1,173,486/5,996,213	Increments 0/15,925/1,232,469; central direct retention $-1,050$; unmeasured quality/time benefits; stress uses a different unrestricted policy.
Advantage original	64/cell; 78+13 weeks	Reference/favorable Beta 378,064/1,286,443	Quality 6/64 and 54/64; retention $-1,755/-6,861$. Separate corrected version admits none.

The separate post-results correction to *The Advantage That Survives* is not a replacement for its original findings. Its forecasting defects and fresh corrected supplement are identified in the source audit. Similarly, we do not infer durability of a selected paid-work policy from another study's unrestricted stress grid. Our selected controller is held fixed in all its own stresses.

3 A venture with an executable transformation

The venture is a **certified numerical-reformulation foundry**. A customer supplies a bounded public or synthetic resource-allocation specification. The service returns an executable representation or experiment, a purpose bridge to the customer's original requirement, explicit numerical assumptions and a checkable result. The primary family changes sparse or low-rank representations. Limited contrasts construct discriminating certificate experiments and complementary scenario/recourse protocols. This is demanding algorithm and engineering work, not clerical output or an idea list.

The term "certified" refers to the specified acceptance contract. Neither an uncertain simulated validator nor the floating-point witness proves every semantic property of a real customer's mission. Mature ordinary tools already solve many such problems: HiGHS supports LP, MIP and QP; SCIP exposes exact-solving and certificate facilities under documented qualifications; independent optimization certificates have a substantial literature [10, 11, 12]. Beta is allowed those capabilities.

Jobs have known scope multipliers 0.65, 1.00 or 1.45, continuous difficulty from 0.25 to 0.95, declared acceptance requirements and one-week or two-week deadlines. The family mixture is 52% representation, 33% certificate and 15% recourse. Scope scales project costs, team references and requirements consistently; a per-project saving is never multiplied by an unrelated weekly workload. All three families are already offerable by Beta. Additional accepted recourse work therefore measures capacity

Table 3: The seven-stage journey is tied to concrete artifacts and decisions.

Stage	Required productive change or evidence
Identify	State a binding formulation, conditioning, coupling or verification limit and a bounded customer objective.
Out-Learn	Establish what ordinary tools and the recipient already solve; record assumptions and dependency versions.
Out-Think	Expose structure that may avoid a dense representation or reveal a discriminating failure mode.
Out-Design	Produce the parameterized executable representation, checker or complementary recourse protocol.
Out-Strategise	Choose offered scope, price, reserve, task acceptance and whether prospective retention is worth its added cost.
Out-Execute	Run the protocol, check the declared contract, correct or abstain, deliver and account for every activity.
Out-Evolve	Test fresh recipient tasks and changed assumptions; retain, revalidate, supersede, retire or revoke according to current evidence.

or acceptance, not a newly invented offering class.

3.1 An executed ordinary-tool witness

A separate numerical experiment makes the transformation concrete while protecting a competent baseline. It solves

$$\min_x \ \tfrac{1}{2}x^\top Ax - b^\top x \quad \text{subject to } Ex = f, \qquad A = D + UU^\top, \tag{4}$$

with positive diagonal D. A Woodbury representation avoids full dense factorization within scope:

$$A^{-1}v = D^{-1}v - D^{-1}U(I + U^\top D^{-1}U)^{-1}U^\top D^{-1}v. \tag{5}$$

A Schur complement $S = EA^{-1}E^\top$ supplies the complementary equality-recourse component. The representation, residual experiment and composition are useful artifacts; neither mathematical identity is an invention claim.

The witness contains 40 fresh in-scope instances and three eight-instance holdouts: dimension/rank, conditioning and a changed tridiagonal base. Every instance has fresh coefficients and eight query pairs. Dense SciPy, an independently written fresh structured Beta and a retained structured template receive the same information, arithmetic precision and within-job factor reuse. The latter two reconstruct numerical factors for every fresh problem. All 192 method/instance records pass the declared acceptance check. Both structured paths take a conventional fallback on the eight changed-base instances.

Table 4: Executed numerical witness. Work is a declared arithmetic-size proxy, not hardware FLOPs or paid-job cost.

Measure	Dense SciPy	Fresh structured Beta	Retained template
Fresh instances accepted	64/64	64/64	64/64
Total counted work	135,171,436	11,720,438	11,720,438
Median solve time, ms	0.1622	0.0582	0.0598
Changed-base fallbacks	—	8	8

The checker independently multiplies the original structure, verifies scaled stationarity and feasibility against a 5×10^{-10} threshold, and checks a positive base-matrix bound. Maximum relative solution difference is 1.4571×10^{-15}; maximum scaled stationarity and feasibility residuals are 2.8998×10^{-17} and 4.9946×10^{-17}. Eight stale-assumption answers and one corrupted answer fail. A partial representation alone fails eight equality-constrained composition fixtures; the complete bundle passes all eight. Competent Beta also has that bundle.

These checks use floating-point arithmetic, not outward-rounded formal proof. The timings exclude instance generation and checking, are local diagnostics, and do not calibrate model latency or economics. Construction time for a human or model was not measured. The retained template ties structured Beta in acceptance and counted work, so this witness establishes *no incremental retained advantage*. It demonstrates what a meaningful transformation and a fair counterfactual look like.

4 Dated calibration and explicit assumptions

We verified official model and pricing documentation and recorded its evidence date, 26 October 2026 in Toronto. The modeled reasoning route is `gpt-6-astra`, `reasoning.mode=pro`, high effort. The execution route is `gpt-6.1-terra`, standard mode, high effort, for visual interpretation and bounded computer use when its acceptance contract can be met efficiently [5, 6, 7]. No customer-job calls to either route were made in this experiment. Product descriptions do not empirically establish the task-success parameters below.

Table 5: Published cost/limit anchors and modeled route choices. USD per million uncached tokens.

Field	Astra	Terra
Input / output price	10 / 50	1 / 5
Published context / maximum output	1,050,000 / 128,000	1,050,000 / 128,000
Modeled per-request input / output cap	200,000 / 25,000	200,000 / 25,000
Published Build tier RPM / TPM	5,000 / 1,000,000	5,000 / 1,000,000

Output budgets include reasoning work. Pro aggregates additional model work at the ordinary rates; we do not invent a fixed Pro multiplier. The long-context threshold is request-specific: above 272,000 input tokens, documented multipliers are 2× input and 1.5× output. The declared decomposition keeps individual requests below that threshold, while complete projects contain many calls. The implemented ledger prices aggregate project token budgets; it does not expand individual API requests or enforce their per-request caps. Weekly token capacity is enforced, but there is no executed RPM scheduler or request-level billing trace. All token costs are uncached; no unearned cache discount appears. Model aliases were documented, but immutable provider snapshot IDs and actual account entitlements were not established. Tool prices inform the calibration register; aggregate tool allowances and operating overhead remain design assumptions [8].

The human-service comparator uses the WCTS GSA schedule effective 18 September 2026–17 September 2027: Operations Research Analyst $85.37/hour, Senior $117.27, Lead $137.64, Project Manager $122.33 [9]. These are billable labor references, not signed customer quotes, willingness to pay or evidence that every mission has an exact human equivalent. We specify team hours and vary their price multiplier. Scope, uncertainty and currency remain visible.

There is no arbitrary expert-use or superhuman bonus. Expertise is represented by the specified decomposition, ordinary tools, metered work, checking and repair process. The source-linked assumption register distinguishes official facts, design assumptions and executed results. No task-matched

Table 6: Nominal unit-scope project design. Token totals include modeled expert work; they are assumptions, not measurements.

Field	Representation	Certificate	Recourse
Human hours: analyst / senior / lead / PM	32 / 16 / 8 / 4	24 / 32 / 12 / 4	48 / 48 / 24 / 8
Astra input / output, million	16 / 2.0	21 / 2.8	36 / 4.6
Terra input / output, million	8 / 0.8	12 / 1.2	20 / 2.0
Additional tool allowance, USD	34	44	74
Production / validator hours	13 / 3	17 / 4.5	25 / 6
Construction share of production	0.55	0.50	0.60

empirical distribution of Astra/Terra quality, token use, turnaround, candidate usefulness, transfer or demand was obtained.

Three profiles span a cautious, reference and opportunity setting. Their names describe model inputs, not estimated market frequencies. Respectively, mean weekly lead arrivals are 1.7/4.5/7.0; base defect probabilities 0.16/0.095/0.05; repair probabilities 0.60/0.76/0.87; conditional validator false-pass parameters 0.14/0.075/0.04; false rejections 0.055/0.025/0.015. Candidate generation rates are 0.24/0.32/0.42, transfer parameters 0.42/0.67/0.82, and weekly unexpected-obsolescence parameters 0.035/0.020/0.012. Costs scale by 1.60/1.00/0.75 and human references by 0.80/1.00/1.15. These joint scenarios are supplemented by independent parameter variation; they are not three calibrated forecasts.

5 Causal world and information boundaries

Four agents—`atlas`, `lyra`, `orion` and `vega` under the modeled `agent.agi.eth` namespace—begin with heterogeneous ordinary experience. Their recipient overlap increases through completed work in every arm and through common Beta improvement. Private graphs are separate per creator: an inaccessible peer capability cannot be counted as already available to a private recipient. The network graph provides common availability to every current authenticated fixture member, without requiring invocation.

Stateless keyed randomness identifies exogenous events by world, job, week and semantic event. A different number of candidate assessments never shifts the next unrelated market draw. Markets, scopes, preferences and relevant production shocks are matched across policies. Endogenous queues, agent assignments and available capabilities can differ; those differences are part of the policy effect, not exogenous privileges. All arms share price options, reserve choices, task ceilings, resource budgets and initial capital.

5.1 Learning has causes and can fail

An accepted paid artifact may produce no candidate. Otherwise, novelty and a modularity draw generate duplicate, narrow, transferable, complementary or misleading candidates. The configured conditional mixture is 20% duplicate, 25% narrow, 29% transferable, 16% component and 10% misleading; incorrect source jobs yield misleading candidates. Components outside a usable recourse bundle have no standalone saving. Only a rights-cleared artifact can enter assessment. A candidate is not made useful because the enterprise has cash to invest.

Hidden world state includes actual correctness, technical transfer, systematic assay bias, unexpected obsolescence and future shocks. Observed state includes current task scope, past started-job counts,

historical scope, current Beta, declared validity, metered costs, observed outcomes and noisy fresh probes. Operating decisions are constructed from the latter. Complete records retain both for audit, but pricing and admission cannot inspect future demand, latent quality or a realized lifetime.

A creator may reconstruct the same method more easily than a peer; narrow methods transfer less well. Recipient overlap reduces the incremental construction saving. Poor transfer can fail to help even after a valid access grant. Misleading candidates can share a systematic assay bias. Thus eight apparently consistent probes need not be eight independent pieces of truth. Research on correlated language-model errors motivates this structure but supplies no task-specific numerical calibration here [13].

5.2 Automated delivery and finite resources

Routine intake, clarification, production, validation, up to two correction attempts and recovery are automated in the model. Each consumes costs and relevant resources. Four workers provide 32 effective production hours each per week; validation has 40 hours; the shared token ceiling is 480 million per week. These effective budgets are model assumptions below advertised ceilings, not verified provider throughput. Intake and retention compete with delivery for the same budgets. Queues and deadlines can make an otherwise correct result commercially unacceptable.

Defect probability depends on known difficulty, recourse scope, ordinary experience and common Beta progress. Weekly failure bursts affect producers and validators; a job-level shared blind spot can raise the false-pass probability across correction attempts. Validators can falsely reject or abstain. Corrections consume additional production, validation and tokens; they may fail or introduce a defect. Similar-model agreement never becomes independent proof.

Beta advances every 13 weeks in the baseline, lowering modeled defects and increasing overlap. Exogenous family/pattern absorption can make a retained capability redundant. The same absorption schedule applies to every arm and persists in current baseline keys, so a new equivalent candidate cannot evade redundancy by receiving a new ID. Faster improvement, changing scope and surprise obsolescence are tested separately.

The runner reserves retention resources before delivery. This is consequential: a campaign can displace high-value work even when its fee is small. The finite policy class does not optimize all possible scheduling orders, asynchronous screening or queue-aware deferral rules. We report this limitation rather than infer that every possible retention architecture is uneconomic.

6 Lifecycle and prospective admission

The reused lifecycle executes candidate, evidence, Alpha assessment, validation and admission states. A candidate remains outside the validated positive graph until its current evidence and incremental-value gate pass. A successful canonical-promotion call cannot bypass admission. Evidence, policy, capability and Beta versions bind every transition. Failure memory is separate from the positive graph.

A campaign requires at least three observed comparable family/pattern starts, sufficient remaining time, cash and resources. The prospective attempt forecast is

$$\widehat{N}_s = \frac{n_{f,k} + 0.5}{w + 9} \min(10, 52 - w, V_s - w)\, q_p\,(0.85), \tag{6}$$

where $n_{f,k}$ counts observed comparable starts, w is the zero-based current week, V_s is declared expiry week, and q_p is one for network availability or one quarter for private routing. The stored elapsed

count is $w+1$, so its eight-week smoothing denominator is $w+9$. Known validity is not the realized remaining useful lifetime. The smoothing, utilization factor and ten-week cap are explicit forecast rules, not empirical estimates.

Eight new recipient/variation probes estimate scope-sized benefit beyond current Beta and recipient overlap. Every observed probe delta subtracts the entire invocation fee. Transfer probability never discounts a fixed attempted-use fee. A one-sided 95% small-sample lower bound and a systematic-risk reserve feed the native gate. With all monetary terms in cents, the gate computes

$$L_s = \overline{\delta}_s - t_{.95}\,\mathrm{SE}(\delta_s) - r_s, \tag{7}$$

$$G_s = \left\lfloor \widehat{N}_s L_s - C_{e,s} - C_{r,s} \right\rfloor, \tag{8}$$

$$G_s \geq \max(1, \lfloor (C_{e,s} + C_{r,s})(h-1) \rfloor), \tag{9}$$

with $L_s > 0$, evaluation cost $C_{e,s}$, remaining cost $C_{r,s}$, tested hurdle h, and the additional correctness, rights, scope, redundancy and freshness requirements. The hurdle changes required net value; it does not relabel actual costs. The exact typed gate and comparison bindings are in the pinned source. A lower confidence bound cannot eliminate common-mode assay bias; out-of-sample use remains necessary.

Self-contained recourse bundles are jointly assessed with representation and certificate components. Independent component admissions cannot silently confer untested joint value. Separate lifecycle tests cover explicit prerequisite versions and downstream invalidation. The economic model does not perform an unrestricted combinatorial search over arbitrary graphs.

On every invocation, access checks current membership, ownership, scope, prerequisites, dependencies, evidence freshness, Beta version, retirement and revocation. Network availability includes every admitted rights-cleared shareable capability; unrelated or confidential customer material and protected evaluation values are excluded. Download, discovery and authorization are not counted as useful reuse. A useful reuse must contribute to a fresh correct, on-time accepted job.

Revalidation pays fresh costs. Supersession, retirement and revocation remove future influence while retaining history. Dependencies cannot revive a revoked component through a historical assessment or composition. Modeled ownership changes, unauthorized readers, stale grants and direct-label boundaries are tested explicitly. Existing Jobs validator roles and settlement remain unchanged. Local AES-GCM and typed SHA-256 Merkle inclusion establish the tested byte relationships; they do not prove truth, profitability, live ENS membership or the production Keccak profile.

The native object-frame boundary is finite. A preserved capacity diagnostic fails at 50 same-owner admissions. The registered study limits total assessments/revalidations to 20 per world and two new campaigns per week, records capacity deferral and makes no arbitrary-scale graph claim. Storage, upkeep, revalidation and validator capacity all have costs.

7 Commercial decisions and disjoint accounting

7.1 Buyer choice and bounded price optimization

A buyer chooses AI, human service, postponement or decline. Given reference price H, latent value V and delay cost d, the implemented utilities are

$$U_{AI} = \widehat{q}V - P - \widehat{\tau}d - \rho V + \epsilon_{AI}, \tag{10}$$

$$U_{H} = 0.985V - H - 3d + \epsilon_{H}, \tag{11}$$

$$U_{postpone} = 0.08V - 0.06H + \epsilon_{P}, \qquad U_{decline} = 0. \tag{12}$$

The buyer takes its best response. Value ranges from 1.05 to 2.30 times the human reference, urgency and tastes vary, and ρ is a profile-specific trust penalty. The seller observes declared scope and its past reliability signal, not latent buyer value. The reliability signal incorporates observed lateness and delayed refund accruals; undetected hidden correctness is unavailable. These response and reputation dynamics are uncalibrated assumptions.

The seller selects a posted-price ratio from 0.55, 0.75 or 0.95 of the scope-matched human reference, rather than imposing a universal discount or cost-plus markup. It also chooses no entry or entry, four/eight weeks of overhead reserve, and a known difficulty ceiling of 0.75/0.95. This is a bounded leader-price/buyer-response model, not a proof of equilibrium in a changing competitive market. The best common commercial controller is selected on Beta and applied to all architectures. We thereby isolate retention and sharing at common freedoms; we do not estimate every architecture's globally optimized maximum.

7.2 A, B and C pay once

The ledger has three disjoint expense categories:

- **A: delivery.** Intake, clarification, paid-job learning and development, production, tools, checking, corrections, hosting and operational governance.
- **B: retention and reuse.** Additional screening, extraction, generalization and fresh assessment; packaging, integration/distribution, catalog discovery, every paid invocation attempt, revalidation and upkeep.
- **C: optional extra investment.** Independent formulation experiments beyond the paid job, only when observed evidence and funds justify them.

Every activity has a unique identity. The paid artifact is not charged a second development bill in B. Customer payment does not erase A. Rejected candidates, failed assays, failed transfer and paid invocation attempts remain expenses. A demonstration's source-job quote less directly attributed A is a direct contribution before shared overhead, not fully allocated job profit.

Seed capital is $30,000 financing. Reservations constrain spendable resources before commitments. Expense accrual creates vendor obligations; invoices create receivables; collections and payments affect cash. Cash cannot go negative. Obligations survive closure. Weekly settlement nets due collections before due payments: this time-resolution assumption may understate intraperiod liquidity needs. The 20-week runoff and maximum six-week job age resolve modeled receivables and refund obligations; none are hidden at the terminal date.

For invoices R, recognized refunds F and complete expense $C_A + C_B + C_C$,

$$\Pi = R - F - C_A - C_B - C_C. \tag{13}$$

Refunds are contra-revenue, not a second delivery expense. The quality-linked revenue measure is

$$R_Q = \sum_{j:\, correct_j \wedge ontime_j} (\text{cashCollected}_j - \text{refundPaid}_j), \qquad \Pi_Q = R_Q - C_A - C_B - C_C. \tag{14}$$

It removes retained receipts from wrong or late outcomes while preserving all costs. A refunded amount is not subtracted again. Hidden correctness is used only by the research evaluator. Positive Π_Q does not imply a satisfactory failure rate.

Ending cash, obligations, reserve and available operating surplus are reported separately. The latter excludes seed, subtracts commitments and reserve, and is capped by reconciled profit. There is no salvage value, token issuance, token appreciation, speculative asset sale or $AGIALPHA utility counted as operating earnings.

8 Freeze, confirmation and uncertainty

The scientific source and protocol were frozen at 02:47:02 UTC on 27 October 2026 before calibrated development. The source commitment is

`0fdc618ffada213cf666a1e66e543dcf7f7fe909726004336665397ccbe2f894`.

Constructed smoke seeds and pre-freeze corrections are preserved separately. They are excluded from discovery and confirmation. Development uses six seeds per profile. Thirteen commercial options (no entry plus $3 \times 2 \times 2$) are scored by mean quality-linked profit minus 0.25 times worst observed quality-linked loss. Private and network retention then each choose off or hurdle 1.0/1.5/2.0 with the same controller. Primary C is off. The selected-policy commitment was recorded before confirmation:

`bf89d815f37f814a22e0b04439f84d6355d3b778de10860b3c3b9b48cd734782`.

Table 7: Nonpooled evidence populations. All enterprise histories use 52 selling and 20 runoff weeks.

Phase	Histories	Design
Bounded development	378	6 seeds/profile; common controller, then retention
Primary confirmation	576	3 profiles × 64 seeds × 3 selected arms
Enabled-assessor challenger	384	Same 64 seeds/profile; private and network
Same-policy stress	3,168	22 conditions × 3 profiles × 16 seeds × 3 arms
Fixed-challenger mechanisms	384	8 conditions × 16 opportunity seeds × 3 arms
Optional-C mechanism	48	Separate 16 seeds; cheap retention; 3 arms
Parameter uncertainty design	144	24 parameter points × 2 seeds × 3 arms
Total	**5,082**	Complete indexed histories retained

Confirmation seed labels are 31000–31063, stress 51000–51015, fixed mechanisms 61000–61015, optional C 62000–62015, and parameter replicates 71000–71001. Fresh task fields and noisy probes use keyed randomness. Changed recourse scope is an explicit family holdout. The separate numerical witness fixes its own dimension/rank, conditioning and changed-base holdouts; it is not pooled with enterprise outcomes.

The prespecified enabled-assessor challenger uses hurdle 1.0, can reject every candidate, and is reported separately from selected retention. Its purpose is to test what assessment and admission actually select, not to replace a primary null result. A cheaper-retention mechanism case scales extra retention costs

to 35%; this is a disclosed parameter change, not a measured cost improvement. Optional extra experiments may defer or decline; no investment count or reinvestment percentage is imposed.

We report means, medians, p05/p95, minima, losses, closures, shortfalls and complete quality compliance. The world-level service gate requires at least 20 accepted jobs, 98% correctness and 95% on-time delivery. Pooled rates do not qualify a world. Monte Carlo intervals use world-level paired differences and conservative small-sample approximations. They quantify stochastic variation conditional on assumptions, not confidence that the assumed market or model performance is correct. The 24-point parameter design is design coverage, not a posterior; fractions of its cells are not probabilities.

9 Results: profitable Beta first

The selected controller enters all three primary markets, retains four weeks of overhead reserve and accepts known difficulty up to 0.95. It prices at 55% of the human reference in the cautious profile and 75% in reference/opportunity. All selected private and network policies turn retention off. Primary outcomes therefore coincide exactly across Beta, private and network arms, including cash, quality and resource use.

Table 8: Primary Beta confirmation. Each column contains 64 independent worlds. All amounts are USD over the same 52+20-week horizon.

Measure	Cautious	Reference	Opportunity
Mean ordinary profit	$121,004	$919,696	$1,810,306
Mean quality-linked profit	$112,543	$897,732	$1,734,211
Median ordinary profit	$121,043	$905,384	$1,866,565
5th percentile ordinary profit	$81,249	$792,400	$1,383,472
95th percentile ordinary profit	$159,716	$1,090,646	$2,091,631
Minimum ordinary profit	$48,649	$729,896	$782,418
Complete quality gate	5/64	44/64	39/64
Loss / closure / cash shortfall	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0

The result preserves profitable useful delivery without requiring retained Alpha. All primary worlds have positive ordinary profit, and none closes or records a cash shortfall. The finding is conditional on scope, demand, reference prices, token budgets, correction costs and buyer response. It is not actual revenue and it is not sufficient evidence of service viability.

Only 5/64 cautious, 44/64 reference and 39/64 opportunity worlds meet the complete quality gate. The opportunity profile's pooled correctness and on-time rates exceed their separate thresholds, yet 25 worlds still fail the joint gate. Positive pooled quality and profit must not erase that failure. Quality-linked profit remains positive because a substantial amount of correct, on-time work is delivered; it does not convert the remaining wrong or late jobs into acceptable service.

10 Results: optional retention and network value

10.1 The selected policy's answer is zero retention

Both private and network selectors choose off in every profile. The primary incremental effects Δ_{private}, Δ_{share} and Δ_{total} are all exactly zero. This is an economically meaningful result within the tested policy class. It does not claim that no useful candidate could ever arise, or that a different screening/scheduling design cannot do better. It says the permitted assessment programs did not

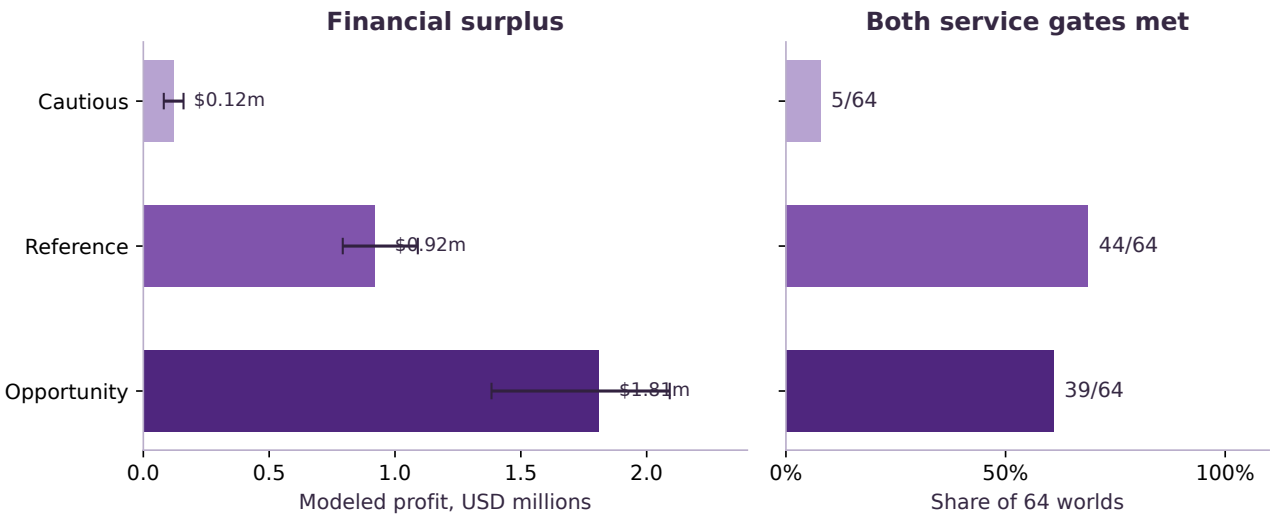


Figure 1: Conditional commercial profit and incomplete service qualification. Profit ranges show empirical p05–p95, not parameter confidence. All selected retention arms equal Beta.

Table 9: Output quality and complete delivery unit costs. Cost ratios divide pooled A expense by pooled output; they are not means of world ratios.

Measure	Cautious	Reference	Opportunity
Mean accepted jobs	70.53	177.27	283.39
Mean correct, on-time jobs	64.11	170.31	268.75
Pooled correctness	94.55%	98.33%	99.50%
Pooled on-time delivery	96.21%	97.70%	95.31%
Pooled A expense / accepted job	$1,669.64	$846.29	$584.25
Pooled A expense / correct, on-time job	$1,836.88	$880.84	$616.07

Table 10: Reconciled mean accounting and cash. Seed financing is USD 30,000; the selected reserve is USD 3,400. All terminal receivables, vendor payables and refund payables are zero.

Mean per world (USD)	Cautious	Reference	Opportunity
Invoices and gross collections	253,198	1,092,151	2,009,595
Refunds recognized and paid	14,432	22,437	33,719
Complete delivery expense A	117,762	150,019	165,570
Additional retention expense B	0	0	0
Optional extra expense C	0	0	0
Ordinary profit	121,004	919,696	1,810,306
Ending cash, including seed	151,004	949,696	1,840,306
Available cash after reserve	147,604	946,296	1,836,906
Available operating surplus	117,604	916,296	1,806,906

justify selection on development data.

10.2 Fresh enabled-assessor confirmation

The prespecified challenger pays for assessment while retaining the option to reject. Its mean profit increments are negative relative to Beta (Table 11). Network-minus-Beta Monte Carlo 95% intervals are $-5,310 to $-4,230 cautious, $-19,139 to $-9,258 reference, and $-173,728 to $-92,173 opportunity. These intervals are conditional model results.

Table 11: Enabled-assessor challenger: mean paired profit differences, USD, 64 worlds/profile. This is a separate diagnostic population, not the selected primary policy.

Profile	Private − Beta	Network − private	Network − Beta
Cautious	-4,199.28	-570.68	-4,769.96
Reference	-14,198.65	0.00	-14,198.65
Opportunity	-132,933.39	-16.89	-132,950.28

The cautious and reference network challengers admit no skills. The opportunity challenger admits four across 64 worlds: three are technically false admissions and subsequently revoked; the remaining one is retired. Eight authorized invocations produce no recorded avoided reconstruction and no useful reuse. Directly linked B charges on those four total $1,029.75, before other rejected candidates and shared discovery. Admission is therefore not profitability by construction.

Many technically useful candidates are rejected. That count is not a measured economic false-rejection rate: a technically helpful method can have too little reuse, too much overlap or excessive retention cost. Prospective uncertainty permits both technical false passes and technically useful rejection. Fully identifying the economic value of every rejected candidate would require a separately designed counterfactual evaluation, not access to hidden future outcomes during admission.

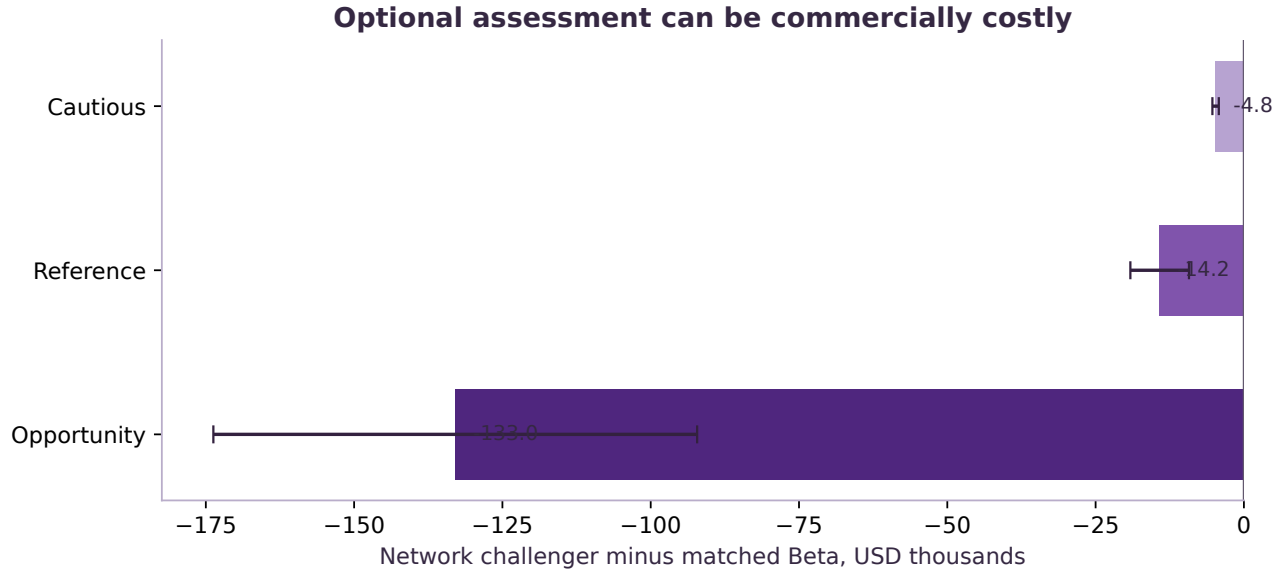


Figure 2: Fresh challenger losses. Intervals describe paired world-level stochastic variation; they do not validate model assumptions.

10.3 Useful reuse emerges when retention is cheaper

The separate fixed-challenger mechanism panel has 16 opportunity worlds per condition. In its cheaper-retention network, 31 capabilities are admitted, 178 attempts are paid for, and 123 useful reuses occur, including 86 by a different agent. Mean B expense is $1,579.51; gross avoided reconstruction is $559.74; the direct retention portfolio balance is $-1,019.77. No world has a positive complete direct balance.

Some individual skills are worth keeping under directly linked accounting. Twenty-one admitted capabilities have positive linked balances before shared catalog discovery. These are upper bounds on fully allocated skill profit, not proof that all retained methods lose money. Their positive contribution is outweighed by unsuccessful screening, assessment and upkeep elsewhere in the program. The study therefore distinguishes a useful individual artifact, an admission program and the complete enterprise.

The cheaper network improves on private retention by a mean $28,395.61, but its paired interval spans $-6,771.05 to $63,562.28. Against Beta it loses $139,214.64, with interval $-243,209.20 to $-35,220.07; only 3/16 enterprise increments are positive. This is not evidence of a robust network advantage or a commercially justified retention program.

Table 12: Fixed-challenger mechanism panel. Useful reuses, direct balance and network-minus-Beta enterprise increment are means per world; money is USD. Each condition has 16 paired worlds with common seeds.

Condition	Useful reuses	Direct balance	Enterprise increment	Quality gate
Base challenger	0.625	-3,739	-167,077	4/16
Cheaper retention	7.688	-1,020	-139,215	6/16
Zero transfer	0.000	-3,720	-170,013	4/16
No turnaround gain	0.625	-3,738	-170,031	4/16
Cost savings only	0.625	-3,738	-170,031	4/16
Faster Beta	0.000	-3,729	-165,866	7/16
Recourse holdout	0.625	-3,739	-187,724	4/16
Assumed quality gain	0.625	-3,739	-167,077	4/16

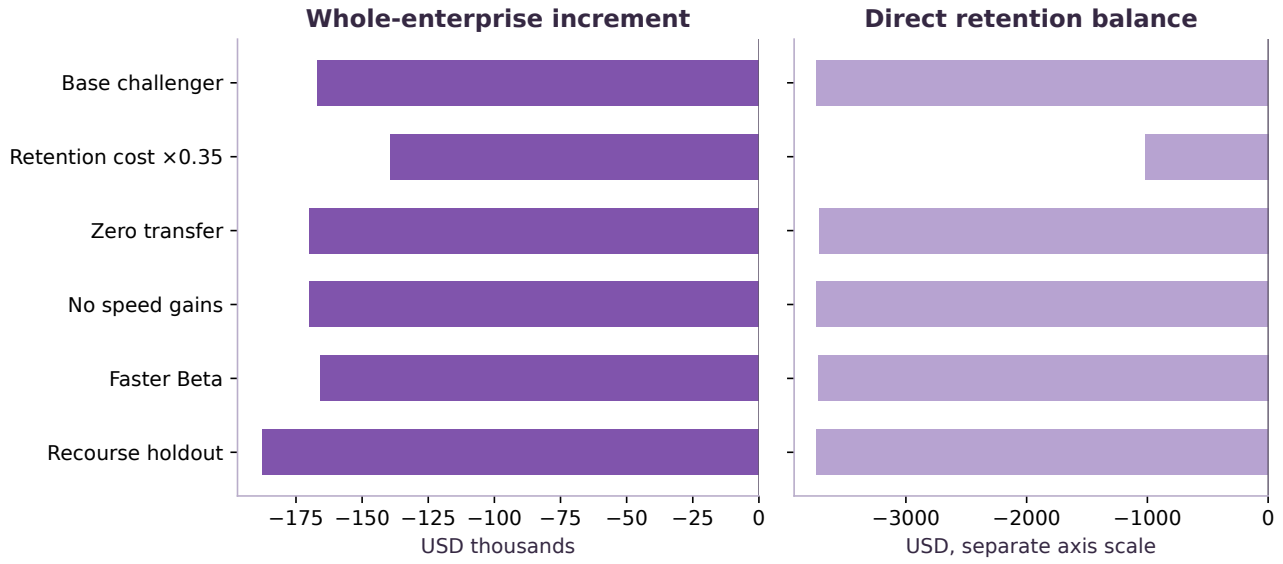


Figure 3: Program balance and enterprise effect are different quantities. Resource savings can be real within the model while the complete program and enterprise increment remain negative.

The direct accounting break-even is transparent. Observed gross reconstruction saving is $50.31 per paid attempt, the fee is $4.20, and mean non-invocation B/C cost is $1,532.79 per world. Holding that observed attempt mix and fixed costs constant,

$$N_{\text{break-even}} = \left\lceil \frac{1532.79}{50.31 - 4.20} \right\rceil = 34 \tag{15}$$

attempts per world, compared with 11.125 observed. This is accounting arithmetic, not a forecast that 34 attempts will arrive or remain useful. More demand can create more screening cost, congestion, overlap and obsolescence; the expression holds those effects fixed. It also does not repay the enterprise's displaced delivery.

The separate optional-C cohort uses different seeds and chooses zero C spending in all 16 cheaper-network worlds. It has 27 admissions and 95 useful reuses, including 65 peer reuses, but direct retention remains negative. Its mean cannot be subtracted from the fixed-panel mean to estimate an investment effect: the populations differ and no extra investment was made.

11 Mechanism, durability and uncertainty

The primary model assumes no retained quality increment and no demand expansion. Removing either is therefore an exact negative control, not proof that those channels have been empirically ruled out. Turnaround benefit is derived from avoided construction work, rather than assigned as an additional expert multiplier. Removing that time saving retains the scoped execution-cost channel while eliminating its resource-release channel.

In the fixed mechanism panel, no-speed and cost-savings-only conditions leave approximately the same small useful-reuse count but reduce the network's mean enterprise increment to about $-170,031. Zero transfer produces no admissions or useful reuse and a mean increment of $-170,013. Faster Beta leaves one admission across 16 worlds and no useful reuse. Changed recourse scope remains negative. Adding an assumed 2.5-percentage-point quality benefit produces the same realized aggregate result in that small base sample; this is low realized exposure, not validation of a quality gain.

The difference between complete enterprise increment and direct retention balance includes displaced work, queueing, corrections, pricing responses, assignment, quality and collections. We do not label that residual a separately identified "capacity effect." The ablations identify interventions within this world model, not additive structural components that can be summed without interaction. Usable capacity, turnaround, recourse acceptance and composition counts are retained in every result record.

Nineteen composed uses occur in the cheaper fixed-challenger network. The witness confirms that its complementary equality component is necessary on its declared fixtures. Neither observation establishes an additional offering unavailable to competent Beta. The no-composition selected-policy stress is correctly null when the selected policy uses no skills; it is not a powered empirical estimate of bundle value.

11.1 The same selected policy under stress

All 22 stresses run the selected paid-work controllers unchanged, on fresh seeds and all three arms. Because retention is off, every retention-specific stress has zero incremental graph effect. Absolute Beta outcomes can still deteriorate. In cautious quarter-demand, mean profit becomes $-1,598.49, quality-linked profit $-4,237.28, 10/16 worlds lose money, and only 1/16 meets quality. Demand uncertainty can reverse the favorable primary commercial result.

Halving capacity produces zero quality-compliant worlds in every profile. Mean profits remain approximately $75,581, $354,296 and $340,124, while pooled on-time rates fall to 86.73%, 84.73% and 71.93%. Profitable accounting therefore cannot be used to hide overloaded service. Payment delays, correlated failures, cheaper human alternatives, doubled compute cost, low recurrence, faster Beta, shorter skill life, expensive revalidation and holdout scope are separately visible in the full stress table and machine-readable evidence.

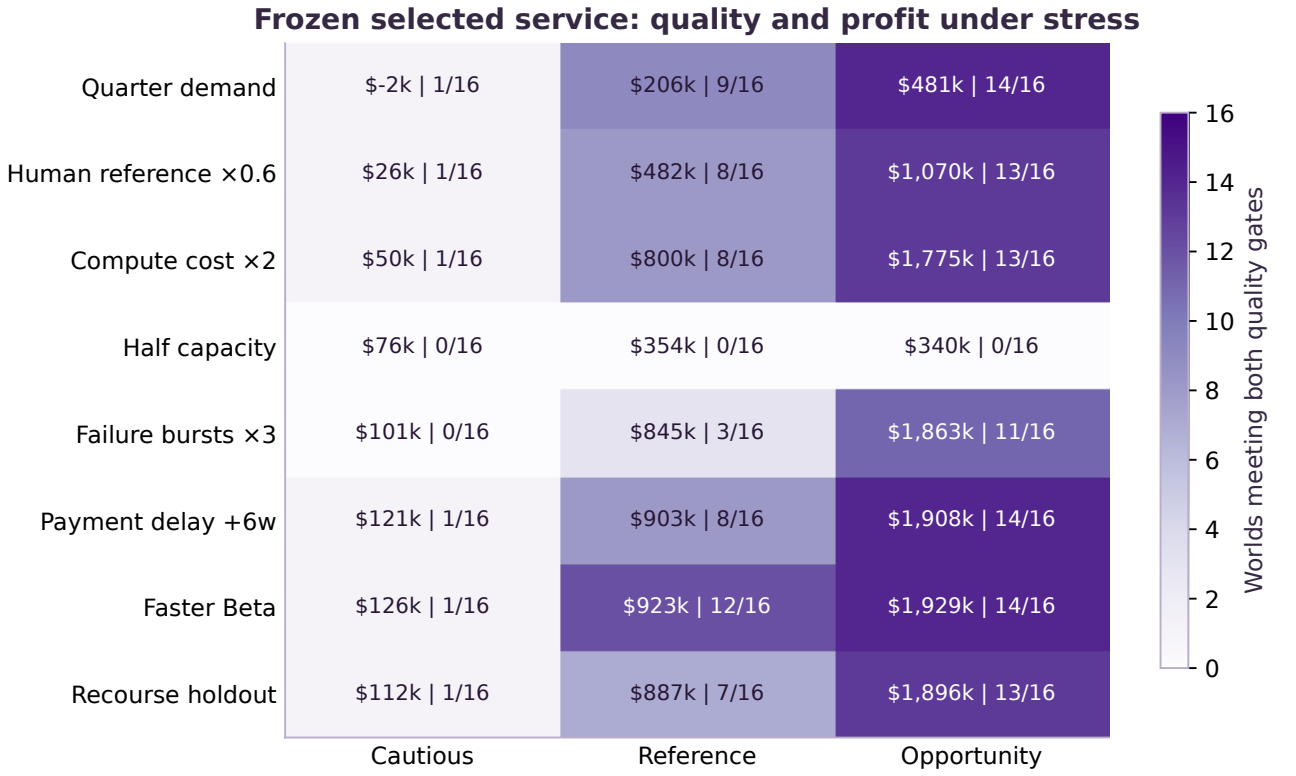


Figure 4: Selected-policy stress examples. Each cell combines ordinary profit and complete quality qualification; the controller is never reselected on the stress.

11.2 Stochastic and parameter uncertainty are separate

The 24-point parameter design changes assumptions jointly and uses two matched shock replicates per point. All 48 Beta histories are profitable, but only 21/48 meet quality. Sixteen points have at least one passing replicate, and only five pass both. These statements describe chosen design coverage. They are not probabilities that a real enterprise will succeed, and two shocks per point are insufficient to estimate each point's tail risk precisely.

12 A trace that can be inspected in English and French

The offline interactive demonstration embeds five exact recorded traces and all principal cohort summaries. Its language switch covers English and French. The visitor can follow one paid mission, its candidate and admission or rejection, a different agent's fresh mission, actual native transitions, modeled costs, complete enterprise finances, failure and faster-Beta cases. It performs no customer API call, identity lookup or transaction. The five stories were chosen after results for illustration; the full cohort statistics and every original history remain available.

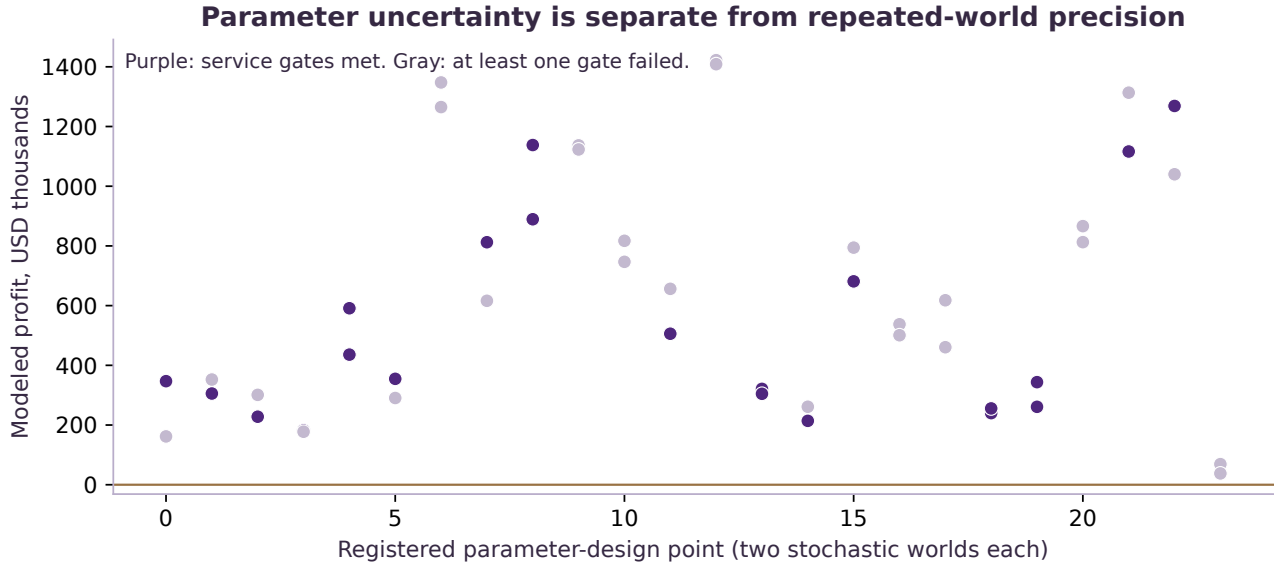


Figure 5: Explicit parameter-design coverage. Positive profits coexist with incomplete quality qualification. The design has no empirical probability weights.

In the cheaper-retention opportunity trace, seed 61000, `atlas.agent.agi.eth` completes certificate job `job-03-05`, priced at $4,452.77 against a human reference of $5,937.03. It is correct, accepted and on time. Its candidate `skill-job-03-05` is assessed through the actual native gate and admitted. Across the recorded world there is one admission, one revalidation, eleven paid attempts, nine useful reuses and seven useful peer reuses. Failed or late attempts remain visible and charged.

That enterprise earns $1,813,143.23 ordinary profit and $1,780,506.95 quality-linked profit, with A $162,395.08, B $1,594.84, C zero and ending cash $1,843,143.23. Its available operating surplus excluding seed and reserve is $1,809,743.23. Yet its direct retention portfolio balance is $-1,187.48, and it earns $5,087.17 less than matched Beta. The demo displays both losses beside the positive enterprise totals. It demonstrates useful learning without pretending that usefulness pays for the whole program.

The HTML rebuilds byte-identically and passes DOM, language-parity, source-binding and semantic checks. A real-browser visual rendering test was not completed because a browser binary was unavailable and the official download returned an invalid response. That limitation is recorded in its QA receipt; no browser inspection is claimed. The PDFs were separately compiled and rendered for visual inspection.

13 Audit, replay and research integrity

Every one of the 5,082 complete histories is stored as an indexed compressed member with byte offsets, raw-record SHA-256, source identity and policy identity. The independent auditor reconstructs 10,583,710 ledger events, 1,075,563 modeled job records and 74,453 native events. It performs 157,283,569 invariant assertions with zero failures. Repeated assertions strengthen record coverage; they are not 157 million independent empirical observations.

Independent artifact checks cover all member hashes, metrics, receipts and source/policy bindings. A separate selector recomputes 63 candidate scores in nine selection groups. A separate report extractor checks 111,515 aggregate quantities against full records. Eleven constructed auditor controls reject

wrong cash, premature settlement, over-refunds, seed counted as profit, duplicate commitments, unauthorized reuse and resource overclaims. The original native controls pass 41/41, the bridge 24/24, and Python causal/ledger controls 16/16.

Nine chosen histories spanning all phases—including admission, useful reuse and a stress loss—replay as identical complete canonical raw bytes, including native projections. These are replay checks, not nine new statistical worlds. All 22 frozen files and eleven vendor identities remain unchanged. Exact hashes, commands, environment versions and receipts are in the evidence companion. The software freeze, policy freeze and phase receipts bind the scientific results; publication, figures and reporting tools are identified separately.

Pre-freeze diagnostic corrections, the native capacity failure, numerical-witness development version, reporting-only metadata repair and browser-download failure are preserved. A reporting extractor initially confused world seed with seed financing; its failed diagnostic and original script remain, and the corrected extractor is independently checked. Scientific outcomes were not replaced. Disk-pressure relocation and interrupted scratch checkout records likewise remain provenance, not hidden experimental retries.

This work does not run the entire upstream release qualification suite, a live provider billing campaign, mainnet ENS, customer settlement, independent market review or deployed autonomous operations. It requires no purchase, new credential, confidential-client data, security change or execution bypass. Public coordination comments register the research; no publication deployment is inferred from them.

14 What the evidence supports

Profitable Beta survives as a conditional commercial result. The source-priced reference model can deliver substantial correct, on-time output at modeled cost below human-team references. A no-retention enterprise is a legitimate outcome. Quality failures and stress losses materially limit the result: product viability requires reliable service, not only positive accounting.

Paid work can produce something useful to another agent in the model. The cheaper-retention panel records fresh peer reuse with explicit scope, fees, recipient overlap and lifecycle transitions. Some individual capabilities have positive directly linked balances. The executed numerical witness makes the transformation concrete while showing that ordinary knowledge may already supply the same capability. Neither demonstration is a live model-learning trial.

The tested acquisition program does not earn its additional cost. Candidate search, rejected assessments, upkeep and resource competition matter. A graph can contain useful methods while the enterprise does better without its retention program. A policy that chooses off protects profitable simplicity; it does not erase those useful methods or establish their universal worthlessness.

Network value remains unproven. Sharing creates availability, not demand. In the cheaper diagnostic the network-private mean is positive but uncertain, and both retention architectures lose against Beta. Ownership, inclusion proofs and authenticated availability are necessary boundaries, not economic mechanisms by themselves. There is no evidence here of permanent scarcity, automatic exponential compounding or self-acceleration from greater volume.

The most consequential limitations are empirical. Model success, token work, transfer, recurrence, customer value, trust and reference project hours remain assumptions. Provider aliases and limits are documented, but account throughput and immutable model snapshots are not measured. The six-seed development search is deliberately small and the scheduling class restricted. The simple buyer

population does not endogenize competition or macroeconomic adaptation. Weekly cash settlement can hide intraperiod liquidity demand. Automated semantic validation can share blind spots. Economic false rejection of untried candidates is not fully identified. The numerical witness measures neither reconstruction labor nor frontier discovery. These limits are explicit conditions of the conclusion, not erased by narrow Monte Carlo intervals or successful code tests.

A positive successor claim would require task-matched evidence that changes these conditions: measured reconstruction savings against a competent recipient baseline; repeat opportunities arriving while the skill remains fresh; accurate acceptance under independent executable checks; and a retention program that covers both unsuccessful search and foregone delivery. Any future empirical pilot should freeze those contracts and prices before fresh customers or tasks, preserve refusal and failure, and maintain the same A/B/C ledger. This paper prepares the executable evidence framework; it does not claim that such a pilot has occurred.

Research conclusion. Paid delivery and retained Alpha are separate sources of value. In this experiment, profitable Beta is preserved, unnecessary retention is declined, useful peer reuse is visible, and the complete added cost remains visible with it. The demonstrated flywheel is a disciplined decision process: deliver, test, keep only what earns its place, and retire it when the baseline or mission changes. A positive, durable economic network flywheel is not established.

A Complete selected-policy stress summary

Table 13: All 22 stresses: the same selected policy. Cells show mean profit in USD thousands; quality-compliant worlds out of 16; loss worlds out of 16. Identical private/network results follow from selected retention off.

Condition	Cautious	Reference	Opportunity
Assumed quality gain	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Capacity half	75.6; 0; 1	354.3; 0; 0	340.1; 0; 0
Cheap retention	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Compute double	49.5; 1; 0	800.4; 8; 0	1,774.7; 13; 0
Constant beta	110.7; 0; 0	889.9; 7; 0	1,877.6; 12; 0
Correlated failure	100.7; 0; 0	845.2; 3; 0	1,862.9; 11; 0
Cost savings only	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Demand half	40.7; 1; 0	443.4; 11; 0	1,007.9; 16; 0
Demand quarter	-1.6; 1; 10	205.6; 9; 0	481.2; 14; 0
Family holdout	112.3; 1; 0	887.2; 7; 0	1,895.8; 13; 0
Fast beta	125.7; 1; 0	923.3; 12; 0	1,928.9; 14; 0
Half transfer	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Human price down	25.5; 1; 0	482.2; 8; 0	1,069.6; 13; 0
Low reuse	117.6; 1; 0	895.9; 8; 0	1,907.4; 13; 0
No composition	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
No demand expansion	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
No quality	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
No speed	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Payment delay	120.9; 1; 0	902.9; 8; 0	1,907.8; 14; 0
Retention triple	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Short life	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0
Zero transfer	119.6; 1; 0	903.3; 8; 0	1,896.4; 13; 0

The semicolon-separated cell entries use the same 16-world denominator. For example, cautious quarter-demand has negative mean profit, one qualified world and ten loss worlds. Every individual median, quantile, cash/closure count, paired contrast and complete history is available in the evidence package; this table is a compact publication view.

B Availability and exact identities

The deliverable contains the executable simulator, pinned native dependency, tests, source-linked calibration and assumption register, frozen selection, all compressed ledgers, audit receipts, numerical witness, English/French HTML, generated figures, LaTeX sources and three compiled PDFs. The complete ledgers are distributed separately from the compact source/evidence companion and restore under `results/`. No production integration or publication deployment is necessary to replay the study.

The outer `SHA256SUMS.txt` binds the final compiled PDFs and archives without a circular self-hash. The evidence companion binds scientific source and raw evidence in detail. A PDF cannot meaningfully contain its own final SHA-256; its exact compiled identity is therefore recorded externally. The archive includes the original compiled bytes alongside the LaTeX build configuration. Source facts remain linked to their dated public documents, while simulation assumptions remain explicitly labeled.

References

[1] MONTREAL.AI. *Transformational Intelligence: Learning to Change the Game*. Supplied public foundation; inspected Alpha commit 51d5ecb3493b01f718713f6873f2aa4e44fe3cce. https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/51d5ecb3493b01f718713f6873f2aa4e44fe3cce/docs/Transformational_Intelligence.pdf.

[2] MONTREAL.AI. *Autonomous Enterprises: SUCCESSOR Omega*; *SUCCESSOR Architecture*. Same inspected Alpha snapshot. https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/51d5ecb3493b01f718713f6873f2aa4e44fe3cce/docs/Autonomous_Enterprises_SUCCESSOR_Omega.pdf.

[3] MONTREAL.AI. *PROOF LOOPS*, research paper v2, author edition. GoalOS commit 1e9a5675b1662e98074fb1c71cf2c5b94c7d259f. https://github.com/MontrealAI/goalos-agialpha-sovereign-machine-economy/blob/1e9a5675b1662e98074fb1c71cf2c5b94c7d259f/docs/paper/PROOF_LOOPS_research_paper_v2_author_edition.pdf.

[4] MONTREAL.AI. *Merkle Roots for the Validated Skill Graph*, V114. Same inspected GoalOS snapshot. https://github.com/MontrealAI/goalos-agialpha-sovereign-machine-economy/blob/1e9a5675b1662e98074fb1c71cf2c5b94c7d259f/docs/paper/Merkle_Roots_for_the_Validated_Skill_Graph_v114.pdf.

[5] OpenAI. *GPT-6 Astra model documentation*. Evidence date 26 October 2026. https://developers.openai.com/api/docs/models/gpt-6-astra.

[6] OpenAI. *GPT-6.1 Terra model documentation*. Evidence date 26 October 2026. https://developers.openai.com/api/docs/models/gpt-6.1-terra.

[7] OpenAI. *Reasoning models and modes*. Evidence date 26 October 2026. https://developers.openai.com/api/docs/guides/reasoning.

[8] OpenAI. *API pricing*. Evidence date 26 October 2026. https://developers.openai.com/api/docs/pricing.

[9] WCTS / U.S. General Services Administration. Multiple Award Schedule 47QRAA19D00CM, rates effective 18 September 2026–17 September 2027. Pages 6–7 and 27. https://www.gsaadvantage.gov/ref_text/47QRAA19D00CM/0ZNBIJ.3VD0D7_47QRAA19D00CM_MASWCTSOY1.PDF.

[10] HiGHS project. *HiGHS optimization software*. Primary project documentation. https://highs.dev/.

[11] SCIP Optimization Suite. *Exact solving mode*. Primary project documentation; configuration and certificate qualifications apply. https://www.scipopt.org/doc/html/EXACT.php.

[12] K. K. H. Cheung, A. Gleixner and D. E. Steffy. *Verifying Integer Programming Results*. arXiv:1611.08832, 2016, revised 2019. https://arxiv.org/abs/1611.08832.

[13] E. Kim, A. Garg, K. Peng and N. Garg. *Correlated Errors in Large Language Models*. ICML 2025; arXiv:2506.07962. https://arxiv.org/abs/2506.07962. Numerical benchmark rates are not imported into this study.

[14] MONTREAL.AI. Research coordination and prior-study provenance, issue #4850; PRs #4950, #4951 and #4952. Exact source descriptions, historical cohort qualifications and inspected commits are preserved in evidence/source_audit.md. https://github.com/MontrealAI/AGI-Alpha-Agent-v0/issues/4850.