AGI ALPHA FLYWHEEL

THE NETWORK THAT LEARNS TO INVENT

Profitable Work, Experience-Trained Specialists and Shared Alpha

Vincent Boucher
President of MONTREAL.AI

An executable scientific relay, with its economic limits intact.
A thermal-design job leaves behind a numerical component. Bounded practice changes its stored representation. Fresh diffusion and inverse-design missions reveal what transfers, what ordinary methods already supply, and whether any additional advantage pays for itself.

TRANSFORMATIONAL INTELLIGENCE ∞

LEARNING TO CHANGE THE GAME

FIND YOUR ALPHA. BUILD YOUR SUCCESSOR Ω.

Evidence scope: locally executed numerical learning and research lifecycle replay; simulated enterprise operations; no customer sales or live authenticated deployment.

Abstract

An engineering team needs a reusable temperature-response calculation; a second team needs fresh diffusion responses; a third needs informative sensors for a finite inverse problem. We connect these useful computational objectives through an inspectable numerical component, then test whether its benefits survive ordinary technical alternatives and complete incremental costs. A researcher-constructed learning system retains an orthonormal basis from twelve founder solves and enriches it through thirty-two residual-directed Gym actions. Four candidates are preserved, one version is frozen, and unseen cases compare the original component, its developed successor and competent numerical baselines. Fresh confirmation comprises 128 paired in-domain cases, 64 shifted cases and 32 inverse-design missions per composition arm. The resulting component reduces mean raw relative field error 26.3-fold on the fresh transfer distribution; this is a measured representation change, not autonomous algorithm invention. Exact correction, assessment, setup and serving remain separately accounted for. All corrected forward methods satisfy the declared numerical requirements. Distribution shift increases the learned solver's correction cost, while the sensor-design gain is also available to the equally funded competent generalist.

A matched, finite-cash enterprise experiment uses sourced service-rate anchors and explicit assumptions about frontier execution, demand, quality and collections. The reference Beta-only policy produces mean modeled operating profit of $191,886 over 104 weeks; no cash revenue was earned by this study. Optional retention decisions reject the numerical candidates at the declared service scale. The contribution is an executable separation of learning, useful transfer, composition and commercial value, including negative evidence and an existing production admission boundary. It does not establish specialist superintelligence, graph superiority, profitable sharing or self-acceleration.

The evidence in one sentence

Experience changed a reusable numerical representation and improved its fresh approximation; complete-cost admission rejected the reference retention program. Complementary experimental design helped, but did not establish a unique network advantage.

1 Begin with the useful work

The immediate work product is a small scientific-software package that computes and checks responses of a heterogeneous diffusion model. Its founder mission represents a thermal-design service: provide reproducible response fields and a numerical envelope over specified material and source variations. A recipient then needs responses across a wider diffusion–reaction regime, and an experimental-design mission needs sensor locations that identify unknown coefficients from noisy observations. The deliverables are executable computations with finite, published acceptance rules. They are not verified physical products, customer engagements or human-expert trials.

This choice makes learned capability tangible. A saved numerical matrix contains information extracted from solved cases. Loading the matrix changes the recipient's calculation, even though the recipient never observes the original training solutions. A second procedure chooses informative sensor locations. These objects can be independently hashed, transferred, executed, rejected and retired. Their practical value is measurable without assigning an agent a specialist label.

The research question is whether such experience can produce an advantage beyond competent ordinary tools, and whether that advantage justifies the additional resources required to develop and share it. The commercial foundation remains a separate question: can expertly operated frontier AI produce accepted digital outcomes at a cost below a price referenced to equivalent human services?

A negative answer about this particular retention program does not refute profitable Beta delivery. Conversely, attractive modeled delivery margins do not establish that learning, a graph or an agent network adds value.

Three distinctions organize the paper. First, a technical result is not a customer outcome. The computation is executed; model-operated engineering and market transactions are simulated. Second, an accuracy improvement is not automatically an improvement in accepted throughput: a competent solver may already meet the requirement cheaply. Third, an accessible capability is not necessarily a capability that every recipient independently possesses. Executable transfer, trained-component transfer and specialist-service access preserve different things.

1.1 A focused portfolio with intact purposes

Offering	Beneficiary and complete computational deliverable	Obstacle, acceptance and recurrence
Thermal response engine	Engineering R&D team comparing proposed material/source configurations; runnable solver, fields, envelope receipt and verification.	Heterogeneous coefficients and repeated solves. Preserve the stated operator, boundary and forcing. Repeat demand is related design revisions, assumed commercially.
Diffusion response service	Scientific-software team evaluating a broader diffusion–reaction model; fresh response fields and checks.	New coefficient/source ranges. Meet the same final residual and field-error tolerances. No right to discard hard cases or replace the target model.
Inverse experiment design	Experimental team selecting a finite sensor plan; selected locations, observations, inferred coefficient pair and uncertainty conditions.	Limited sensors, noise and non-identifiability. Recover the correct member of a published sixteen-element coefficient grid; retain failed reconstructions.

Table 1: Research offering scopes. Human-team scope equivalence and recurring demand are assumptions. The numerical artifacts and their stated finite tests are executed.

The work has a transformational aspect when it changes the representation or the experiment: reuse a low-dimensional response space instead of repeatedly starting from an uninformative state, or choose a measurement that resolves uncertainty instead of collecting more uninformative data. These are established scientific methods. The novelty claimed here is their joined, reproducible learning–transfer–composition experiment and its accounting, not the invention of reduced-basis methods or optimal design.

The finite discretization is deliberately reproducible on an ordinary CPU. It is much smaller than many industrial simulations. A positive result cannot be extrapolated to industrial geometry, physical fidelity or expert service quality without another evaluation. This limitation makes the empirical claim narrower; it does not turn the result into an assumed financial uplift.

2 The actual system and its boundaries

The implementation baseline is the repository commit `80b345828f34fd8a6e592977dd17ef37141305f4`. Coordination was refreshed through issue #4850, including general-execution owner #4960 at `2163c2e12c1a7385241aca9bf813815b40c6e82c` and the active specialist-network lane. This research is

additive under `studies/agi-alpha-invention-network-v1/`. It does not replace a registry, executor, publisher or existing study.

The protected production capsule on the frozen main branch supports a bounded public-source-selection scope. A generic diffusion component cannot truthfully be inserted into that scope. Its production admission therefore remains **NOT QUALIFIED**. The general-execution draft is evidence of ongoing engineering, not evidence that arbitrary specialist capsules, model accounts or authenticated peer trials are already qualified.

For the research replay, eleven unchanged files from the existing `beyond-beta-network-v1` lifecycle at commit `2727262caa52766a78916f2ed81f5afaf668edce` are reused with recorded Git-blob and SHA-256 identities. The actual native Journal, TrustRegistry and CostLedger are imported from the frozen main implementation. The research adapter distinguishes simulated mission authority from execution, retains rights and version bindings, and keeps real research costs unknown where no invoice or metering record exists. Signing a fixture does not authenticate an ENS owner on mainnet.

The supplied Transformational Intelligence, Validated Skills Graph, Proof Loops, Gym and Successor Ω materials establish the intended progression: useful work, retained evidence, bounded development and fresh outcome checks. Their architectural proposals are not counted as executed performance. The source audit records the reviewed originals, exact identities and applicable interfaces. A valid provenance chain establishes which object was used; it does not establish that the object is useful.

2.1 Why earlier negative retention results matter

Earlier research editions provide distinct counterexamples. The Profitable Work First study reported attractive conditional Beta economics in its reference world while retaining no optional capability in its primary campaigns. A separately frozen diagnostic improved one network policy slightly but still underperformed Beta after complete cost. Other editions recorded assets that were reused without repaying their own incremental costs. The audit preserves their source links and distinguishes reported results from new replications.

Those failures suggest three bottlenecks: screening can cost money without creating an admissible asset; expected reuse can be too small or mismatched to the admission belief; and a capability can improve a task without saving enough of the enterprise's scarce resource. The present experiment does not repair those findings by assigning guaranteed gains or lowering costs until a win appears. It constructs a real changed artifact, compares against a strong technical baseline, and asks which bottleneck remains.

3 The inspectable relay

3.1 Agent A: the founder mission

The founder solves twelve thermal cases. Each case specifies diffusion coefficients, a localized heterogeneous component, a reaction term and a spatial source. The complete response fields are retained as an orthonormal basis using two-pass orthogonal enrichment. The initial artifact, v1, contains twelve basis vectors. The modeled founder offering includes a reusable numerical engine, so its solves, checks and basis construction are allocated to A once. The executed founder receipt independently checks the twelve fields and their envelope; it is not customer acceptance of the retained basis. Additional

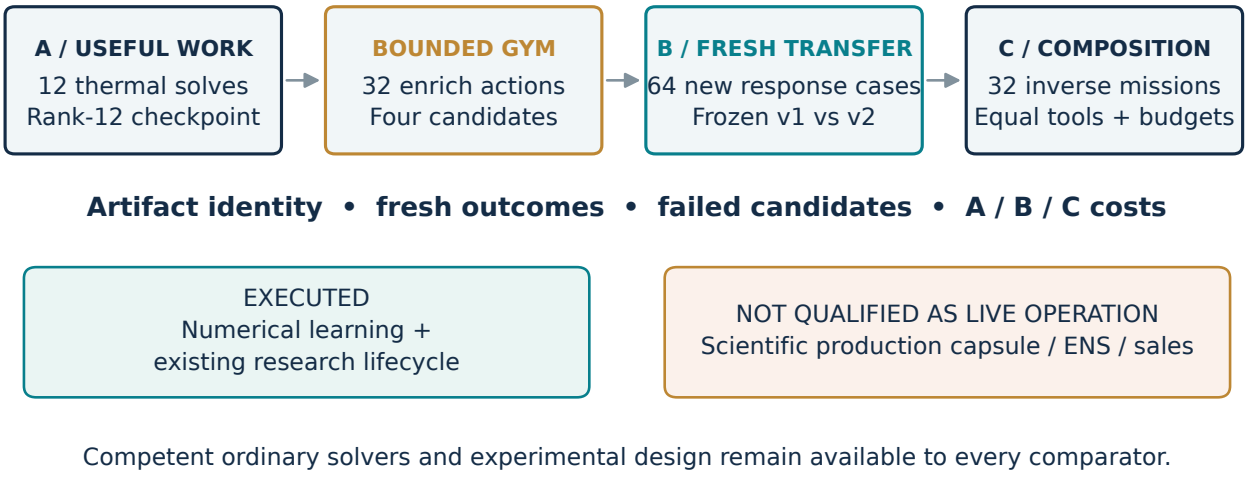


Figure 1: The executed relay and its authority boundary. Founder computations create v1; bounded experience updates a numerical basis to v2; protected fresh work evaluates transfer and composition. Local research artifact use is separate from production capsule admission, which remains unsupported for this scientific scope.

packaging and subsequent lifecycle costs remain separate. No customer paid for the research proxy.

"Agent A" names a scripted research role. The surrounding algorithm, task generators and acceptance rules were constructed by the AI-assisted research team. No frontier model independently invented the basis algorithm during a purchased customer job. Nevertheless, the basis entries are determined by executed solutions, rather than written as a desired performance score. This is identifiable numerical learning.

3.2 Gym: change a representation through bounded experience

The Gym exposes observations, allowed actions, transitions, budgets and termination. Its state includes the current basis, available training cases and remaining enrichment actions. An observation evaluates residuals for a declared pool of seventy-two practice cases. An action selects an available practice case, solves it, orthogonalizes the new field and records the resulting artifact. The controller selects the largest residual, subject to a cap of thirty-two enrichment actions. A case cannot be selected after removal; an exhausted budget terminates the campaign.

Four candidate checkpoints at ranks 20, 28, 36 and 44 are assessed on development cases. The frozen selection rule prioritizes final accepted count, raw accepted count, then mean raw field error, with rank as a tie-breaker. It therefore selects for numerical fidelity, not secretly for final enterprise profit. All candidates, observations, assessment work and the original development attempt remain available. The initial six-sensor inverse task reached a development ceiling; a two-sensor, higher-noise challenge was specified before confirmation, with the original control retained.

The update changes a stored numerical representation. It does not modify proprietary model weights, learn a general skill-selection policy or demonstrate that an agent learned how to learn. Sensor selection is an executable procedure applied to current mission fields. Keeping these mechanisms separate prevents a larger basis, a higher practice score or more inference from being mislabeled as a general

intelligence improvement.

3.3 Agent B: use the component on unfamiliar work

Fresh diffusion–reaction cases broaden coefficient and source ranges. The recipient loads the exact frozen artifact and computes a reduced initial solution. A residual-corrected solver then returns a response meeting the same final acceptance criteria as the ordinary methods. The raw approximation is also evaluated separately. Thus the experiment can detect learning even when a correction step makes the final outputs equally accurate.

The transfer domain shares a discretization and operator family with the founder but changes the parameter distribution and job context. It is a meaningful related-domain test, not evidence of unrestricted transfer across physics or software ecosystems. The artifact requires compatible grid indexing, operator parts, units and numerical libraries. Out-of-distribution cases deliberately increase anisotropy, local contrast and source concentration; their failures and overhead remain in the results.

3.4 Agent C: compose numerical and experimental expertise

The composition mission evaluates sixteen possible coefficient pairs. A sensor-design procedure derives low-dimensional response features and greedily selects locations using an information-gain criterion. No hidden true coefficient is available to the design procedure. Observations are generated from a held-out true grid member plus the declared noise; reconstruction minimizes the finite observation mismatch.

The comparison includes the numerical component alone with fixed sensors, the design procedure alone with the ordinary forward solver, and their combination. The primary competent generalist receives the same design method and ordinary tools. Fixed sensors are an ablation, not the generalist ceiling. Every arm receives the same candidate grid, sensor count and noise realization. Measured forward, design and reconstruction costs are attributed to each arm even when measurement reuse avoids physically rerunning identical code in the experiment. Unmetered research orchestration and authoring remain unknown; additional enterprise coordination costs are modeled explicitly.

This design can reveal complementarity without assuming a network advantage. If informative sensors improve reconstruction but the competent generalist can use the same procedure at equal resources, the result supports experimental design and does not establish retained Alpha.

4 Numerical method, acceptance and protection

4.1 The discrete system and the learned object

For parameter vector θ, the finite-difference model is

$$A(\theta)u = b, \qquad A(\theta) = k_x A_x + k_y A_y + c A_c + \rho A_r, \tag{1}$$

on a square with homogeneous Dirichlet boundary conditions. The grid has $44 \times 44 = 1,936$ unknowns. Positive diffusion coefficients, a nonnegative local contrast and a nonnegative reaction coefficient make the operator symmetric positive definite. The public generators define the exact distributions. They are assumptions about a computational workload, not empirical industrial demand.

For an orthonormal basis V, the reduced approximation is

$$\widehat{u} = V\big(V^{\mathsf{T}}A(\theta)V\big)^{-1}V^{\mathsf{T}}b. \tag{2}$$

The projected affine operators are built explicitly. The stored basis preserves learned numerical information; a prose explanation of the method would not reproduce its entries. The construction is a greedy reduced-basis method, not a claim that an arbitrary neural checkpoint has been trained. Established reduced-basis and POD work supplies the methodological context; the sensor procedure uses an SVD of current response fields.[1–4]

4.2 Acceptance includes an independent numerical check

Ordinary baselines include sparse LU, incomplete-LU-preconditioned GMRES and a separable spectral preconditioner with conjugate gradients. Incomplete LU is used with GMRES, not assumed to be an SPD preconditioner for CG. The primary ordinary solver is selected on development runtime and frozen. Valid factorization reuse is allowed whenever matrices repeat. An explicitly declared improved-generalist numerical policy tests a stronger ordinary choice on the shifted regime; this is not a newly measured API-model release.[5, 6]

Final residuals are recomputed by a flux stencil separate from the sparse assembly. A direct solve supplies a research reference for field-error assessment. The final tolerances are relative residual 2×10^{-8} and relative field error 2×10^{-6}. Raw-surrogate tolerances are separately declared and are not substituted for the final requirement.

There is also a computable discrete error bound. Let $r = b - A\widehat{u}$ and let α be a valid lower bound on $\lambda_{\min}(A)$. Positive semidefiniteness of the local-contrast part permits a lower bound from the two directional Laplacians and the minimum reaction term. Then

$$\|u - \widehat{u}\|_2 \le \frac{\|r\|_2}{\alpha} =: E, \quad \frac{\|u - \widehat{u}\|_2}{\|u\|_2} \le \frac{E}{\|\widehat{u}\|_2 - E} \quad \text{when } \|\widehat{u}\|_2 > E. \tag{3}$$

The second relation follows from the reverse triangle inequality. Without a valid stability bound a small residual alone is not a general field-error guarantee. Neither this bound nor the direct reference measures discretization error, real material properties or physical model mismatch.[7]

4.3 Protection and uncertainty

Training, development and fresh confirmation use distinct case seeds. Candidate bytes and source identities are frozen before root-supplied confirmation seeds are evaluated. No post-confirmation candidate tuning is admitted into the primary comparison. The shared research team and public generators do not constitute an external, adversarial evaluation service; the protection is procedural and source-auditable. Public release intentionally makes the cases reproducible after the experiment.

Paired cases receive identical equations, reference answers and mission noise. Numerical methods run with one BLAS thread, warmups and rotating execution order. Wall-clock differences at millisecond scale remain vulnerable to host noise; per-case rows and paired uncertainty are reported alongside iteration counts and error. Timing savings are not generalized to another machine by a clock-rate multiplier.

The reported 95% timing intervals are percentile intervals from 5,000 bootstrap resamples of paired cases (seed 81221) within one local run. They do not measure reproducibility across machines or independent sessions, and resampling does not remove host drift shared by several case pairs.

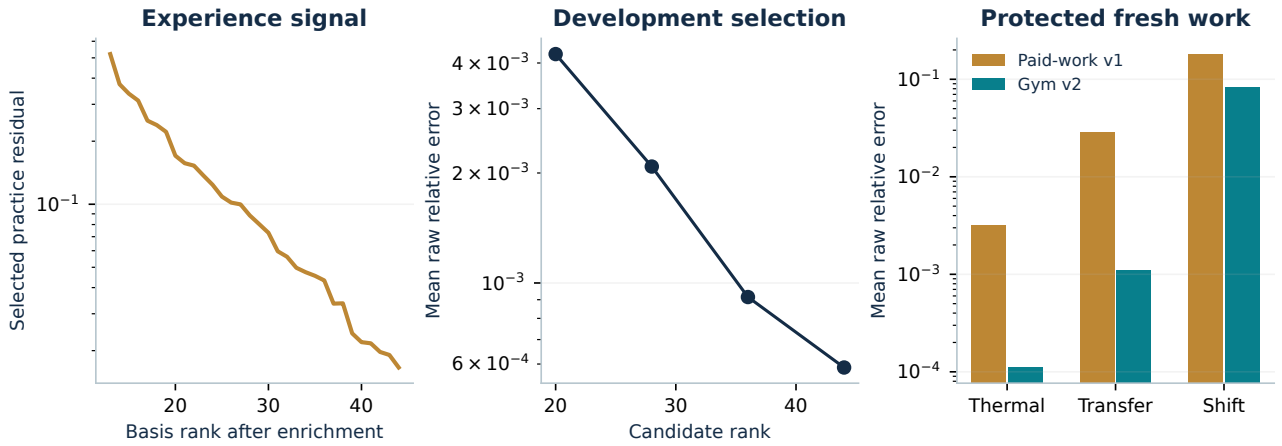


Figure 2: Numerical development and fresh performance. Candidate development results are retained; fresh v1/v2 comparisons use frozen artifacts. Practice residuals and fresh field errors answer different questions and are not plotted as a single interchangeable score.

Confirmation estimates finite-distribution performance, not the probability that the assumed workload resembles a real customer portfolio. Repeating seeds reduces Monte Carlo uncertainty conditional on the generator. It cannot remove uncertainty about the generator, the scope of a human service or market demand.

5 What changed, what transferred, and what did not

The frozen Gym candidate has rank 44, compared with rank 12 for v1. On the 64 fresh transfer cases, mean raw relative field error falls from 0.02869 to 0.00109, a 26.3-fold reduction. However, neither raw component meets the separately declared raw acceptance rule on any of those 64 cases. The corrected methods all pass. Accuracy learning is therefore established within this workload; a standalone approximation meeting the recipient's requirement is not. On the 64 fresh thermal cases, v2 passes the raw rule on 9 cases.

The meaningful learned change is the basis itself. It supplies a better initial representation of previously unseen response fields. This does not imply a proportional reduction in completed-job cost. The recipient's competent baseline can solve the original finite system, and the corrected learned method still pays for projection, a linear solve, residual checks and any remaining iterations.

Method	Cases	Accepted	Time (ms)	Iterations	Raw error
Sparse LU	64	64	4.075	0.00	–
ILU-GMRES	64	64	16.357	42.20	–
Expert baseline	64	64	2.106	15.08	–
Improved ordinary	64	64	2.037	15.08	–
Paid-work v1	64	64	2.121	13.95	0.02869
Gym v2	64	64	1.981	12.19	0.00109

Table 2: Fresh numerical confirmation. Runtime includes assembly, per-case setup, solve and verification. One-time artifact construction and independent assessment are reported separately. A learned checkpoint's raw error is distinct from its corrected accepted output.

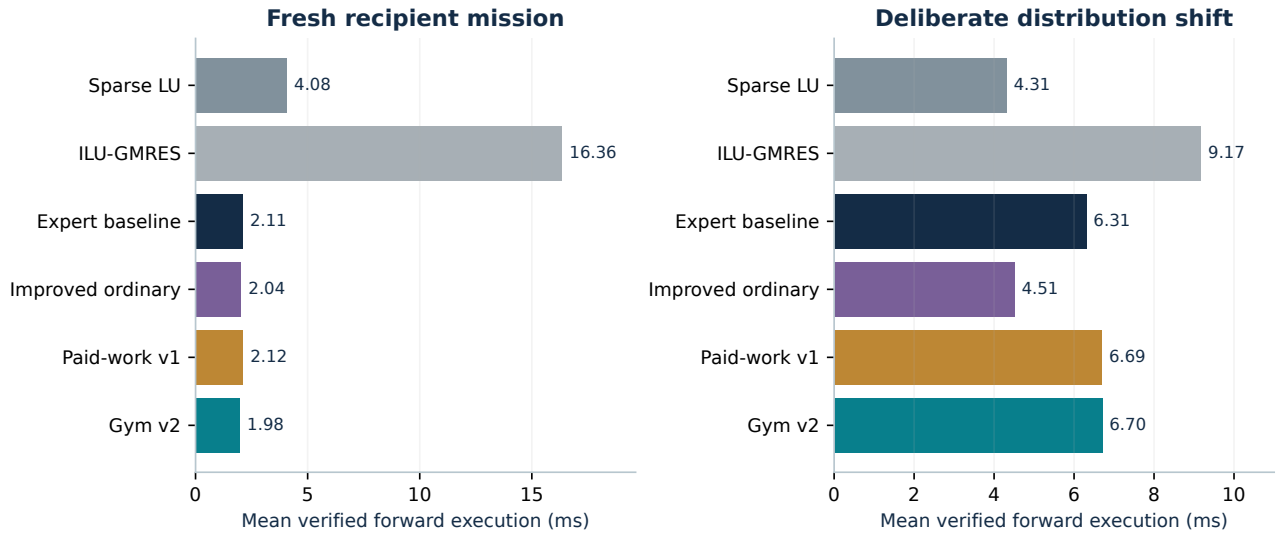


Figure 3: Paired transfer and distribution shift. Ordinary numerical baselines remain available to recipients. A lower raw error can coexist with no runtime gain, and distribution shift can impose negative transfer. All failed cases remain in the denominator.

On fresh transfer, the observed mean saving of v2 against the frozen ordinary baseline is 0.125 ms per solve (paired bootstrap 95% interval 0.052 to 0.199 ms). Against the improved ordinary implementation it is 0.056 ms (interval -0.007 to 0.125 ms). The improved policy uses the same numerical algorithm in-domain, so this difference also exposes local timing noise. On shifted cases, v2 takes 6.702 ms, compared with 6.306 ms for spectral-PCG and 4.508 ms for the improved ordinary router. No shifted raw learned output passes the raw acceptance rule. The negative transfer remains part of the conclusion.

5.1 Composition and the equally funded generalist

	2 sensors, 5% noise		6 sensors, 1.8% noise	
Method	Correct/total	Time (ms)	Correct/total	Time (ms)
Fixed sensors	7/16	43.13	16/16	38.91
Solver only	7/16	38.04	16/16	38.14
Competent generalist	10/16	44.15	16/16	40.16
v1 + design	10/16	40.56	16/16	42.40
v2 + design	10/16	39.00	16/16	39.39

Table 3: Inverse-design composition, with identical candidates, observations and sensor budgets. The competent generalist includes the design procedure. Fixed-sensor arms are ablations. Success concerns the finite numerical identification task only.

With two sensors and 5% peak-field noise, fixed sensors recover 7/16 coefficient pairs. Both the competent generalist and the v2 combination recover 10/16. With six sensors and 1.8% noise, every arm recovers 16/16. The observations support the value of selecting informative measurements in the harder finite task, but the composition does not improve identification over the equally resourced generalist. The sixteen-case challenge is too small to support a broad reliability claim. Combination runtime is reported without converting it into an assumed price or scientific discovery.

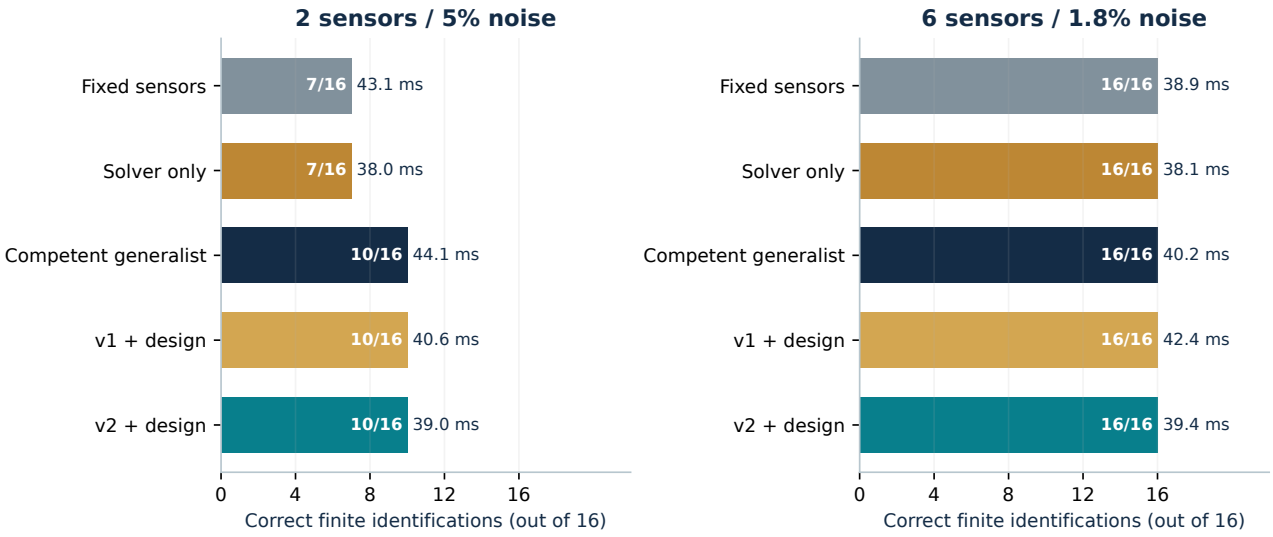


Figure 4: Component ablations and composition. Accuracy is measured on protected inverse cases and shown with measured task execution cost. Any gain shared by the equally funded generalist is ordinary available capability, not exclusive network Alpha.

5.2 The productive frontier at fixed resources

Define the finite empirical productive set at budget B and required acceptance q as

$$\mathcal{F}(B,q) = \{j : \Pr(\text{accepted outcome for } j) \geq q,\ \mathbb{E}[C_j] \leq B\}. \tag{4}$$

The operational definition must fix the beneficiary, original requirements, workload distribution and accounting boundary. A faster approximation that abandons accuracy does not enlarge this set. A composition expands the frontier only if it crosses a required reliability or cost boundary that the equally resourced generalist cannot already cross.

The descriptive time-frontier curves show local differences in verified execution latency. They do not show a new reliable offering beyond the competent generalist: all corrected forward methods meet the numerical requirement, and the composition ties generalist identification. No prospectively specified service-level deadline was crossed uniquely by the learned network. The enterprise model therefore assigns no automatic new-product revenue or quality premium to these objects.

6 From numerical value to enterprise value

6.1 Frontier intelligence is Beta, with verified public specifications

The reference commercial workflow assumes competent use of frontier reasoning, coding, tools, validation and correction. It does not model a one-shot no-tool prompt as the ceiling. Public API documentation checked on 27 October 2026 lists `gpt-6-astra`, `gpt-6.1-sol` and `gpt-6.1-terra`. Their Standard short-context input/output tariffs are respectively \$10/\$50, \$2/\$10 and \$1/\$5 per million tokens. The published roles can be replaced by better verified successors; the model names are not restrictions on future design.[8–10]

The source register records context/output limits, documented reasoning efforts, tool support and

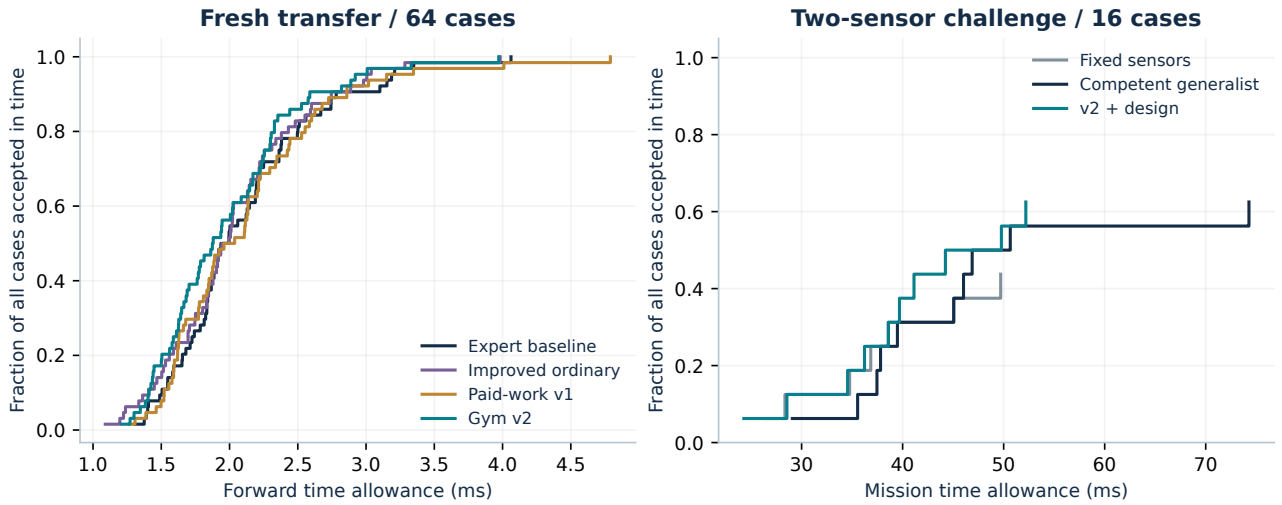


Figure 5: The observed finite resource frontier. The curves count accepted outcomes within a solver/mission time allowance. They do not value a proposed venture, establish physical validation or assume additional demand. The strongest ordinary comparator is included.

advertised rate tiers. Those specifications establish neither account entitlement nor available quota. No API call or account-funded experiment was executed. ChatGPT Pro is not an API model or evidence of API access.[11] All dollar-valued model consumption is an explicit token-budget assumption using the sourced tariffs. Fine-tuning is not assumed; the portable trained object in this experiment is a numerical component.

6.2 The human-service reference and the pricing decision

SDSU lists an external scientific-computing service rate of $80/hour; Vanderbilt research software engineering starts at $84/hour; UTEP lists an external analyst rate of $190.31/hour for services including study design. These are service-rate anchors, not quotes for our artifacts.[12–14]

The central scope estimate uses 72, 96 and 120 hours at $84/hour for the three offerings: $6,048, $8,064 and $10,080. The effort includes formulation, implementation, independent numerical checks, correction, documentation and delivery. Every effort estimate remains an assumption; rates for more complex specialist work can differ. Demand and price acceptance are modeled separately.

A frozen finite grid compares price fractions 0.30, 0.45, 0.60, 0.75, 0.90 and 1.05 of the human-service reference. Development simulations select 75% under the declared reference belief. This is a conditional grid optimum for the stated objective, not a global optimum or an observed willingness to pay. The same selected prices apply across matched policies. The comparison therefore does not give shared specialists easier customers or a hidden price increase.

6.3 Disjoint costs and cash

The model distinguishes three expense categories:

A: delivery. Frontier use, tools, computation, acceptance, corrections, base delivery overhead and the common startup capacity needed to supply the outcomes. Learning required to finish the paid job belongs here once.

Input	Declared reference assumption
Seed / common startup	$20,000; 96 engineering hours at $84, totaling $8,064
Base resources / overhead	40 agent-hours per week; $100 per week
Model-use bundles (thermal / transfer / inverse)	$66 / $96 / $138; itemized assumed token counts at sourced tariffs
Integration / acceptance bundles	$38/$22; $54/$32; $68/$48, respectively
Estimated human participation per attempt	Operator 0.25 / 0.40 / 0.50 hours; reviewer 0.25 / 0.35 / 0.50 hours, included in those bundles
Optional retention / sharing / Gym	$192 assessment and packaging; $42 sharing; $168 Gym development, plus recorded numerical work
Second specialist, if elected	Additional $192 assessment/packaging and separate serving cost
Screen / maintenance / invocation	$2.40; $1/week; $0.002 per specialist invocation
Collections and reserves	Assumed 3% + $0.30 fee; four-week warranty exposure; eight-week base operating reserve

Table 4: Commercial assumptions that materially determine profit. These are neither observed invoices nor a requirement that every automated service use human review. Human participation is an explicit conservative service-design choice in this edition; sensitivities do not establish the real need or sufficiency of that choice.

B: additional retention. Candidate screening, extra assessment (including every rejected rank), packaging, admission, sharing, serving and upkeep that are not already in A.

C: optional development. Additional Gym experience and discretionary development beyond the paid deliverable. Candidate assessment belongs to B; extra Gym enrichment belongs to C. Capacity expansion is separately identified.

Research labor and subscription/model use during this authoring session were not fully metered. Actual complete research cost is therefore *unknown*, not zero. Commercial scenario costs are projected operational expenses. They must not be described as invoices paid in this study.

For a simulated enterprise, cash obeys

$$K_t = K_0 + \text{collections}_t - \text{refunds}_t - A_t - B_t - C_t. \tag{5}$$

Cash-based operating surplus excludes seed capital. Outstanding obligations, receivables, operating reserves and available funds are separately reported. Failed delivery, corrections, late collection, default and refunds are retained. Capacity exclusions represent displaced work; they are not also charged as fictional cash expenditures. There is no token issuance, staking yield or proposed venture valuation in operating revenue.

The model's nano-dollar accounting preserves numerical computation and small serving costs instead of rounding them out of existence. Native CostLedger replay checks a complete selected synthetic history and, separately, preserves unknown actual-research costs. The full ensemble is reconciled by the study's transaction audit; it is not misrepresented as thousands of native production deployments.

7 Matched enterprise experiments

7.1 Five choices, equal starting conditions

The policies are: competent generalists without retention; private retention; private retention with optional Gym; selective sharing; and sharing with optional Gym and composition. They begin with the same seed, opportunity streams, quality shocks, price menu and resource limits. Candidate admission uses development evidence and declared demand beliefs. The policy cannot inspect future demand, future true quality or its exact date of obsolescence.

No policy is forced to buy training or retain an asset. The primary screen first asks whether even free elimination of all numerical computation could repay minimum retention cost; if that optimistic bound fails, the controller declines without consulting an unpurchased Gym result. When forecast incremental value is negative, declining is the correct policy outcome. Composition uses an equally funded generalist with ordinary access to the experimental-design procedure. Because no graph-specific advantage is claimed, an ordinary library with the same authorized bytes and permissions is an equality control, not a deliberately impoverished search baseline.

The horizon is 104 operating weeks plus collection runoff. Three primary environments vary opportunity arrival, trust, willingness to pay, quality, model cost, default and successor timing. A stronger frontier scenario changes base execution cost and quality for every policy, while an independent numerical comparison tests a stronger ordinary solver. The former is simulated; the latter is executed. The assumptions and all sampled histories are included.

First, resources remain fixed. A separate high-demand experiment allows expansion only after available earned funds exceed the capital and reserve conditions and forecast capacity value exceeds expansion cost. Extra output from purchased capacity is labeled expansion. It is not counted as evidence that retained intelligence compounded.

7.2 Results and uncertainty

Beta-only result	Adverse	Reference	Favorable
Collections (USD)	9,934	228,078	759,583
Refunds (USD)	469	3,175	6,282
Complete costs (USD)	19,639	33,016	62,595
Operating profit (USD)	-10,174	191,886	690,706
Useful outcomes	1.86	40.42	132.52
Cost/useful outcome (USD)	10,530	817	472
Closing cash (USD)	9,826	211,886	710,706
Outstanding obligations (USD)	0	0	0
Uncollected receivables (USD)	0	0	0
Reserved cash (USD)	800	800	800
Available cash (USD)	9,026	211,086	709,906
Available earned funds (USD)	327	191,086	689,906

Table 5: Conditional enterprise results, mean over 200 matched worlds per primary environment. All monetary results are simulated USD; the study earned no customer revenue. Closing balances follow the collection-runoff period. Available cash includes remaining seed funds; available earned funds excludes seed capital and is floored at zero separately in each world. Percentile variation describes the chosen model, not uncertainty about its real-world validity.

Mean Beta-only modeled profit is $-10,174, $191,886 and $690,706 across the adverse, reference and

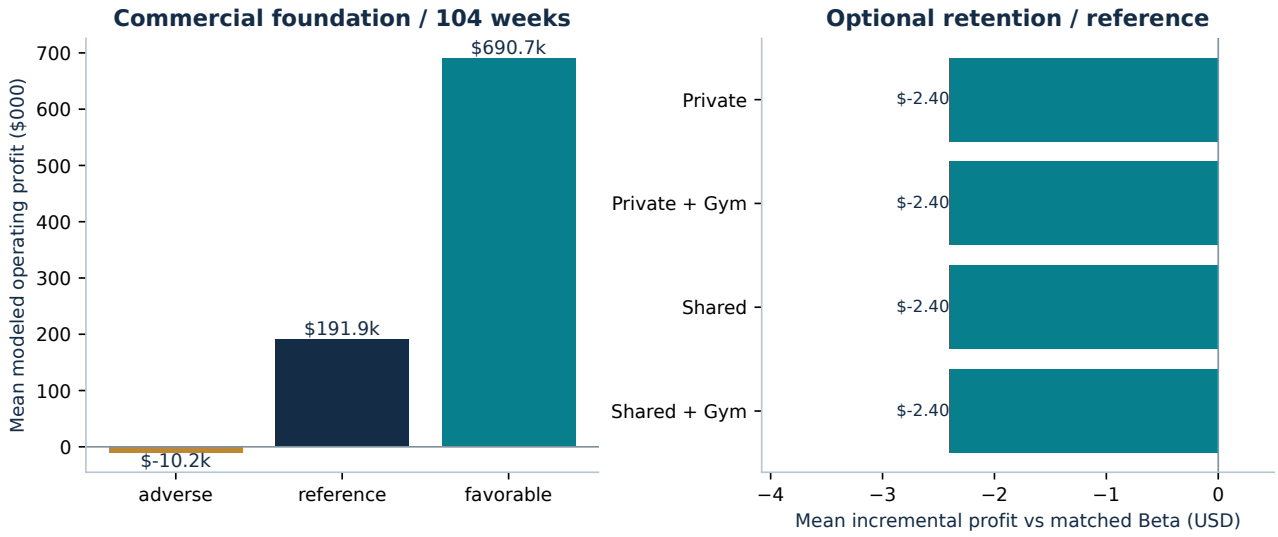


Figure 6: Commercial foundation and incremental retention. Beta profit is conditional on the declared market and delivery assumptions. Incremental Alpha is assessed against matched Beta after all additional costs. A profitable reference enterprise is compatible with rejecting optional retention.

favorable environments. The reference enterprise supplies 40.42 useful outcomes on average at a pooled complete cost of $817 each. Its false-pass share among accepted outcomes is 1.52%, which remains a modeled quality risk. The shared/Gym policy's paired mean increment is $-2.40. No primary policy retains or trains a capability after its optional economic assessment. Differences from Beta arise from screening, rather than from claimed profitable peer reuse.

The compulsory-retention diagnostic is deliberately separate from the selectable main policies. It asks what happens when the numerical artifact is retained despite its economic assessment. It includes extra assessment, packaging, serving, maintenance and any elected development. It cannot be used to infer the outcome of a rational policy that would decline the purchase.

In the forced-retention diagnostic, the shared/Gym/composition policy changes mean profit by $-641.71 relative to its matched Beta. The diagnostic does not rescue the optional program. Additional conditions stress congestion, correlated errors, late collections, low value, absent demand and rapid baseline improvement. The separate surplus-financed expansion condition changes purchased capacity; its output is reported separately and is not attributed to learned intelligence. In the congestion and surplus-expansion diagnostics, the optional policies show about USD61.18 mean advantage (Monte Carlo standard error USD111.01), despite zero retention, training or technical use. The small screening expense changes later cash-admission and order-mix outcomes; it does not debit a training or screening time budget and is not Alpha.

The resulting dollar levels are not forecasts. Unmeasured quality, customer trust and service equivalence dominate uncertainty. A simulation can identify consequences of those assumptions and expose accounting errors; it cannot validate the assumptions by generating more seeds.

8 The admission boundary is an economic inequality

Let Δt be the validated time saving per eligible use against the competent recipient; p_c the applicable compute price per second; f additional serving cost; N eligible uses before retirement; and F all

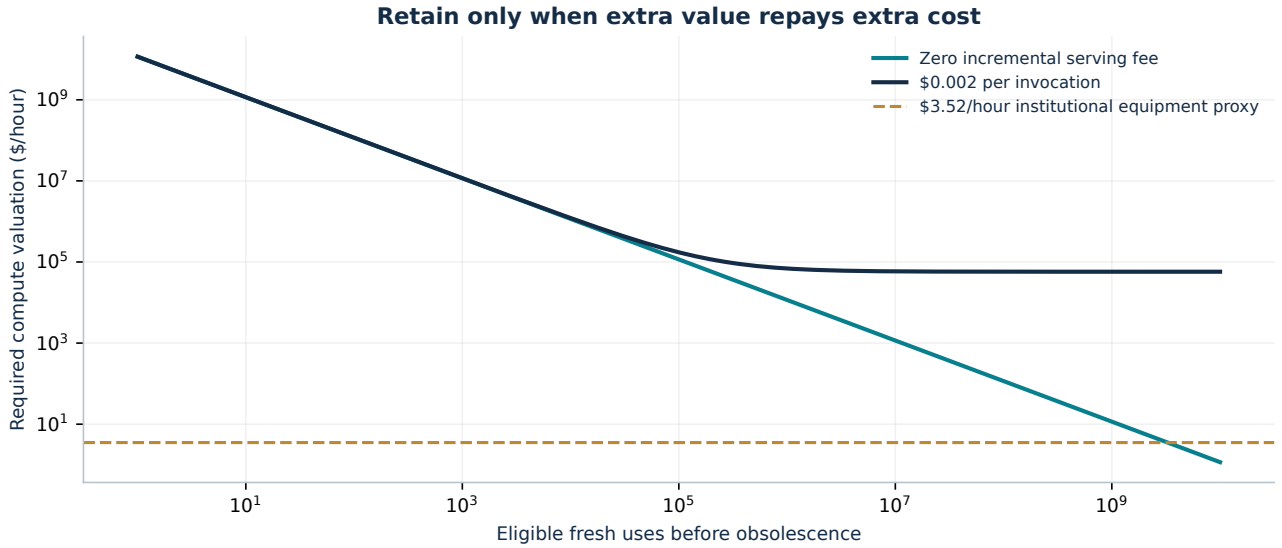


Figure 7: When retention could repay its additional cost. This algebraic boundary uses measured timing at the tested scale and explicit cost assumptions. Regions requiring a different compute valuation or workload are sensitivity scenarios, not demonstrated industrial speedups.

additional assessment, retention, failed-candidate and Gym costs. With maintenance M,

$$\Delta V = N\big(p_c \Delta t - f\big) - F - M. \tag{6}$$

This is a restricted decision model: it values measured runtime only. There is no assigned quality premium, avoided human project or extra demand merely because an artifact exists. If reliable new quality were measured, its customer value would require a separately supported term.

Two necessary conditions follow. First, if $p_c \Delta t \leq f$, more reuse cannot repay a positive fixed investment. Second, if the per-use margin is positive, then $N > (F + M)/(p_c \Delta t - f)$ is required. A large number of agents does not change either condition. It helps only insofar as it creates eligible, authorized and valuable uses after coordination, maintenance and congestion.

At the declared \$3.52/hour equipment proxy, the observed transfer time saving is worth approximately \$0.000000122 per solve. An assumed \$0.002 invocation cost already exceeds that gross benefit, before extra assessment or training. Even with zero serving fee and no maintenance, repaying \$402 of assessment, packaging, sharing and additional Gym development would require approximately 3.28 billion eligible uses at the same measured saving. With the invocation fee, the compute valuation would first need to exceed about \$57,512/hour before any positive fixed-cost repayment is possible. These are algebraic thresholds, not claims that the measured solver advantage survives at a different scale or price.

The \$402 illustration is a lower bound: \$168 assessment, \$24 packaging, \$42 sharing and \$168 Gym development. It excludes screening, initialization, numerical assessment, maintenance, serving and the optional second specialist's additional \$192 assessment and packaging. The enterprise histories include these additional costs when applicable; the illustration is not the full cost of the composite program.

This boundary explains why improving a component can be scientifically worthwhile while admitting it commercially is premature. It also gives a practical next experiment: identify a workload whose bottleneck the learned object actually removes, with enough related demand to repay measurement and upkeep. Increasing a supposed asset value, erasing rejected searches or moving original development

cost into someone else's ledger would not answer that question.

9 Three transfer forms and their consequences

Form	What this edition transfers or exercises	What the evidence does not establish
Executable skill	Reduced solver, residual check and sensor-design procedure; reproducible local code.	Novel algorithm discovery, portable execution in every environment or superior graph routing.
Trained component	Exact numerical basis bytes, versions, dependencies and hashes; local loading changes the response approximation.	Neural fine-tuning, proprietary weights, general intelligence or compatibility outside the declared discretization.
Specialist service	Existing research authority/rights and serving-cost interfaces exercised with synthetic actors; positive transport control and revocation.	Provisioned protected endpoint, live ENS authentication, actual external invocation or possession of the specialist by every caller.

Table 6: Availability, possession and performance are separate claims.

The empirical numerical candidates are evaluated with observed numerical resource costs and explicitly estimated additional B/C scopes. Actual complete research expense remains unknown. If that assay rejects admission, the study preserves the rejection. A separately labeled positive conformance fixture tests byte-preserving transport and revocation with declared synthetic economics. Its recipient executes the first saved transfer case again using the retrieved bytes. This is a later replay, not another fresh trial. Fresh confirmation had used the frozen local artifact directly. The control is not evidence that the empirical candidate earned its way into production or delivered fresh authenticated work after admission.

The source-pinned lifecycle tracks rights, prerequisites, version, baseline, evidence expiry and protected evaluation material. Eligible capabilities can be made visible to every authorized fixture member; private rights restrict them to the owner. Revocation and baseline changes invalidate use under the expected rules. The production boundary remains in force. Existing Jobs validators and settlement are neither modified nor emulated as actual on-chain acceptance.

10 Interpretation: what has been established

Experience-based learning is a productive research direction. Silver and Sutton's essay and Ineffable Intelligence's stated mission motivate investigating agents that learn through environmental interaction. They do not imply affiliation, equivalent methods or reproduced outcomes here.[15, 16] FunSearch and AlphaEvolve offer precedents for evaluator-guided executable improvement; their accomplishments do not transfer to a new system by analogy.[17, 18]

This study's strongest result is methodological and experimental: the learned object, fresh work, ordinary alternative, transfer mechanism and economic boundary can all be inspected together. It rejects the tempting inference that every improvement is automatically an asset. It also preserves the

Claim level	Finding and limit
Profitable Beta delivery	Positive in some declared enterprise worlds, negative in adverse conditions; entirely simulated commercial evidence.
Technical learning	Executed change to a numerical representation, with fresh error comparison and retained training history.
Useful peer transfer	Exact component used on related fresh numerical work; local research roles, without live network admission.
Positive retention or Gym net value	Evaluated against complete additional costs; not established at the reference service scale.
Positive network-sharing value	Not established by transport conformance or accessibility.
Expanding productive advantage	No claim of sustained expansion beyond the equally resourced competent generalist.
Specialist ASI	An ambition; no human-expert comparison or general superiority established.

Table 7: The claim ladder prevents one successful step from standing in for the whole mechanism.

positive commercial hypothesis that expertly operated frontier models may supply useful services profitably without such assets.

10.1 Limits that determine the next study

The workload is synthetic, low-dimensional in its parameters and restricted to a fixed grid and affine operator family. It lacks human trials, customer purchases, physical validation and measured frontier agent performance. The confirmation is protected procedurally within one AI-assisted research session, rather than administered by an independent laboratory. Timing is local. The enterprise assumes task quality and customer demand. The generic scientific component is outside the frozen production capsule scope.

These limits identify concrete next tests. The first is a customer-relevant scientific-software task with a costly repeated computational bottleneck and independently specified acceptance. The second is a competent frontier generalist that may discover or construct the same ordinary method, with all model and tool costs recorded. The third is an actual authorized service or component transfer with versioned operating permissions and a fresh recipient. Only then should an expanding network be credited with earned economic advantage.

11 Reproduction, authorship and publication

The companion contains the complete code, original and final development records, frozen candidates, case manifests, confirmation outputs, all enterprise histories, source registers, native-interface identities, tests, figures and an offline English/French demonstration. The README provides exact commands. Frozen measurement outputs are not overwritten by a reproduction; new results receive a new directory. Timings may differ, while numerical identities and tolerances are directly checkable.

The paper is attributed as requested to Vincent Boucher, President of MONTREAL.AI. AI assistance was used extensively for research, implementation, literature retrieval, simulation, analysis, figures and drafting. The numerical algorithms and environments were researcher constructed; the executed training updates the numerical basis. No customer endorsement, independent institutional validation,

proprietary model training or autonomous invention is represented as having occurred.

This edition follows the existing repository review and publication path. The current release lease, protected checks and prior editions are preserved. A review artifact or source link is not a deployed service. The companion records the final publication state and verified links without treating an unmerged draft as a qualified live release.

Research decision. Preserve profitable useful delivery as the commercial foundation. Preserve the learned numerical component as evidence. Admit additional capability only when fresh useful performance and complete incremental value justify it. At the tested scale, a better representation is real; a profitable shared specialist network remains unestablished.

References

[1] Gianluigi Rozza, D. B. P. Huynh, and Anthony T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):229–275, 2008. doi: 10.1007/s11831-008-9019-9. URL https://iris.sissa.it/handle/20.500.11767/12327. Accessed 27 October 2026.

[2] Carmen Gräßle, Michael Hinze, and Stefan Volkwein. Model Order Reduction by Proper Orthogonal Decomposition, 2020. URL https://arxiv.org/abs/1906.05188. Accessed 27 October 2026.

[3] Michael Hinze and Martin Kunkel. Residual Based Sampling in POD Model Order Reduction of Drift-Diffusion Equations in Parametrized Electrical Networks, 2010. URL https://arxiv.org/abs/1003.0551. Accessed 27 October 2026.

[4] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-Driven Sparse Sensor Placement for Reconstruction. *IEEE Control Systems Magazine*, 38:63–86, 2018. doi: 10.1109/MCS.2018.2810460. URL https://arxiv.org/abs/1701.07569. Accessed 27 October 2026.

[5] SciPy community. scipy.sparse.linalg.cg, 2026. URL https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.linalg.cg.html. Accessed 27 October 2026.

[6] SciPy community. scipy.sparse.linalg.splu, 2026. URL https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.linalg.splu.html. Accessed 27 October 2026.

[7] pyMOR developers and contributors. pymor.reductors.coercive, 2026. URL https://docs.pymor.org/2026-1-1/autoapi/pymor/reductors/coercive/index.html. Accessed 27 October 2026.

[8] OpenAI. GPT-6 Astra Model, 2026. URL https://developers.openai.com/api/docs/models/gpt-6-astra. Accessed 27 October 2026.

[9] OpenAI. GPT-6.1 Sol Model, 2026. URL https://developers.openai.com/api/docs/models/gpt-6.1-sol. Accessed 27 October 2026.

[10] OpenAI. GPT-6.1 Terra Model, 2026. URL https://developers.openai.com/api/docs/models/gpt-6.1-terra. Accessed 27 October 2026.

[11] OpenAI. Pricing: ChatGPT Work and Codex, 2026. URL https://learn.chatgpt.com/docs/pricing. Accessed 27 October 2026.

[12] San Diego State University. Computational Science Research Center: Services, 2026. URL https://www.csrc.sdsu.edu/services/. Accessed 27 October 2026.

[13] Vanderbilt University ACCRE. Research Software Engineering Service, 2026. URL https://www.vanderbilt.edu/accre/research-software-engineering-service/. Accessed 27 October 2026.

[14] University of Texas at El Paso. Data Analytics Laboratory: Services, 2026. URL https://www.utep.edu/science/data-analytics-lab/services.html. Accessed 27 October 2026.

[15] David Silver and Richard S. Sutton. Welcome to the Era of Experience, 2025. URL https://storage.googleapis.com/deepmind-media/Era-of-Experience%20/The%20Era%20of%20Experience%20Paper.pdf. Accessed 27 October 2026.

[16] David Silver / Ineffable Intelligence. Ineffable Intelligence: Mission and A Note from Dave, 2026. URL https://www.ineffable.ai/. Accessed 27 October 2026.

[17] Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli, and Alhussein Fawzi. Mathematical discoveries from program search with large language models. *Nature*, 625:468–475, 2024. doi: 10.1038/s41586-023-06924-6. URL https://www.nature.com/articles/s41586-023-06924-6. Accessed 27 October 2026.

[18] Alexander Novikov, Ngân Vũ, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog. AlphaEvolve: A coding agent for scientific and algorithmic discovery, 2025. URL https://arxiv.org/abs/2506.13131. Accessed 27 October 2026.