

THE POSSIBILITY FOUNDRY

Autonomous Enterprises That
Invent Their Next Capability

From Autonomous Capability Synthesis
to Measurable Expansion of Economic Opportunity

SUCCESSOR Ω

Powered by TRANSFORMATIONAL INTELLIGENCE ∞

Don't just ask what intelligence can do.
Ask what it can build that changes the answer.

Research direction: Vincent Boucher
23 October 2026 · Executable digital laboratory

Bounded synthesis. Simulated physics. Measured failures.
Applicable \$AGIALPHA utility is discussed; none is activated in this study.

Abstract

Can an enterprise construct an enabling capability that makes a family of previously unaffordable investigations feasible? We study a bounded version of this question in an executable, synthetic spectroscopy laboratory. The enterprise receives fixed scientific requests, an initial 64-channel scanner, and a disclosed construction grammar. It acquires 24 noisy reference spectra, evaluates 60 combinations of measurement design, observation count, and calibration, then emits and executes a reusable instrument. The selected capability combines 16 data-dependent point measurements, affine calibration, and a learned linear decoder. Its 20-exposure operating cost becomes 26.056 exposures per single-session mission after allocating all 15,504 discovery and verification exposures across the frozen campaign.

At the declared budgets, the capability makes 1,678 of 2,048 missions feasible, compared with zero for the original scanner. Total campaign exposures fall from 174,080 to 66,704, a 61.68% reduction. Transfer is partial: the unchanged instrument passes 204 of 512 withheld paired-change investigations. An independently authored construction policy and an expert-derived analytical software comparator discover the identical instrument with lower search costs; the proposed exhaustive policy therefore does not win the builder comparison. Calibration removal reduces feasible missions to 518; a mismatched decoder reduces them to zero. Unseen narrow spectral lines defeat the instrument on all 128 stress cases.

The emitted first capability also acquires the observations used to construct a temporal rate-identification capability. Search selects a three-time-point, offset-exponential estimator, which passes all 128 fresh temporal tasks. Those tasks are affordable conditional on an existing parent capability, but not when the entire parent discovery cost is charged to this small successor campaign. The experiment executes through the released AGI ALPHA workbench and an unchanged Jobs draft compiler. Native spectroscopy admission through Manager/Fleet, the held newer Harness integration, contract execution, physical validation, external human-expert comparison, and the \$100 → \$10,000 commercial target remain unestablished. The contribution is an auditable construction-and-evaluation mechanism and a measured change in a defined resource frontier, not a demonstration of open-ended invention or realized profit.

The result to remember

A constructed instrument changes the affordable work set. A competent alternative builder can construct the same instrument more cheaply. Both findings are part of the result.

Evidence at a glance

Executed contribution	60 instrument proposals; one emitted measurement program; four frozen mission families; 12 successor proposals; a separately emitted temporal program.
Frontier change	0 → 1,678 feasible missions under the fixed resource and quality requirements.
Strong comparison	Both shared-infrastructure builders tie scientific outcomes and beat the foundry's total campaign cost.
Transfer boundary	Withheld paired-change success: 204/512. Narrow-line stress: 0/128.
Successor boundary	128/128 marginally affordable; 0/128 affordable from scratch after the full ancestor charge.
Economic evidence	No settled proceeds or executable commercial offers are established; complete USD and human costs are unknown.

1 The productive object is a possibility

An enterprise ordinarily receives a task and searches for an answer. Here, it receives a portfolio for which its existing means of investigation are too expensive. The decision variable changes: it must choose what reusable means to construct before returning to the portfolio.

The central question is therefore:

Can an enterprise discover what it needs to become, construct that missing capability, and make an entire family of previously impractical missions worth pursuing?

This paper gives an executed, deliberately narrow answer. Within a supplied digital laboratory and construction grammar, a program selects and constructs a lower-cost measurement interface. It then demonstrates which fixed requests cross an affordability threshold, which remain outside it, and what happens when the constructed capability is removed.

The experiment separates three propositions that are easily conflated. First, a system can select a capability rather than receive its final design. Second, that capability can change a portfolio's feasible work set. Third, the particular search policy can outperform other capable builders. Our evidence supports the first two within the declared scope and rejects the third against the comparators used here.

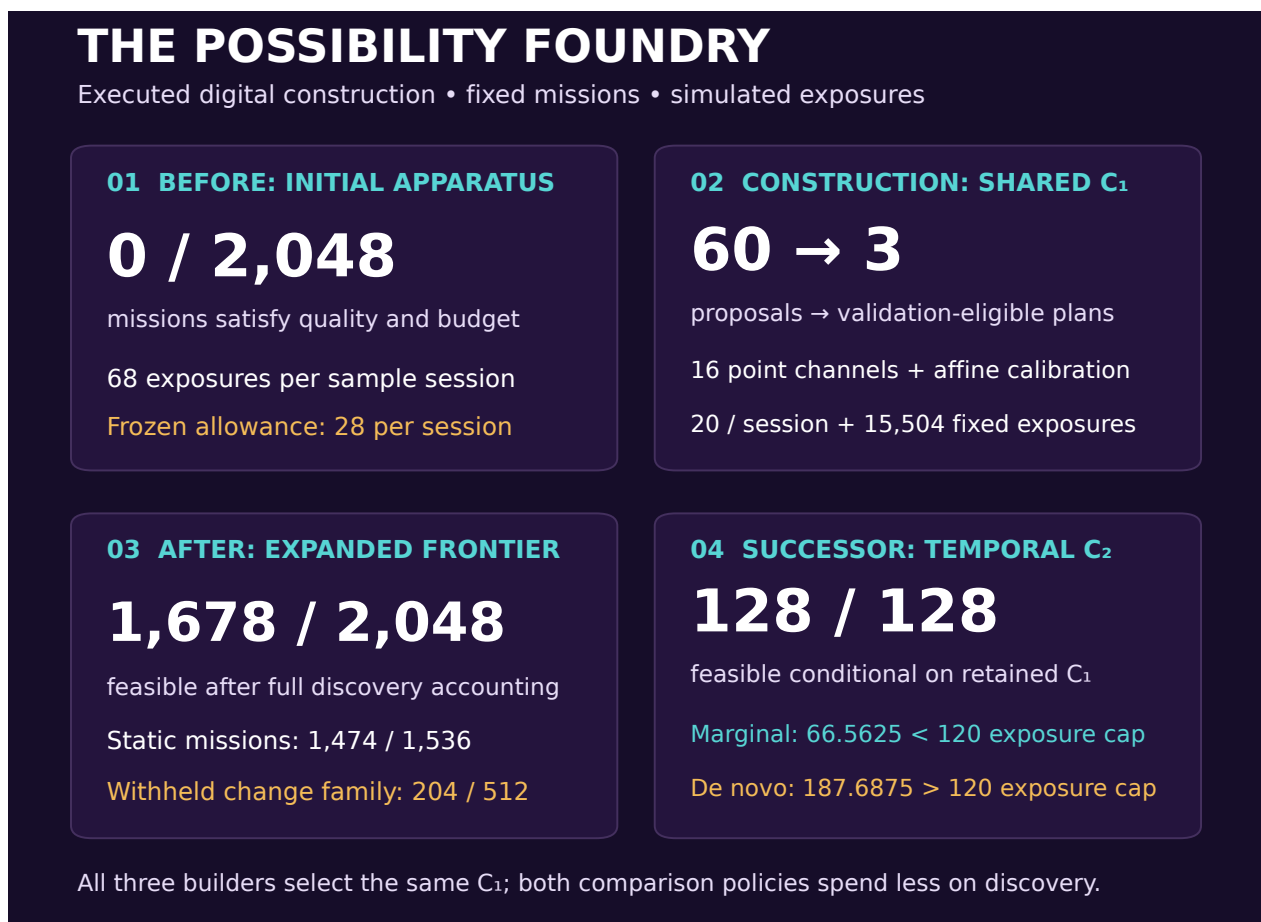


Figure 1: The possibility cycle. The initial scanner meets many quality checks but exceeds the declared operating limits. A measured search constructs a reusable instrument, expanding the affordable set while retaining failures and all fixed costs. That instrument helps construct a temporal capability. The final panel distinguishes affordable reuse of an existing parent from an unaffordable de-novo successor campaign. Every resource number is a simulated detector-exposure count.

1.1 Relation to the preceding research

Writer #4’s *Transformational Autonomous Enterprises* examines supplied-capability inheritance with fresh checks for different missions (MONTREAL.AI, 2026). The new decision here is the instrument to construct: its measurement locations, associated subspace dimension, calibration mode, and decoder are selected from observations and comparative evaluation. The final instrument and temporal successor are not entered as winning blueprints.

Invent the Economy treats the joint technical and commercial conditions that make a new economic possibility viable (Boucher, 2026). This experiment holds task requirements and resource rules fixed. It does not construct buyers, agreements, demand, prices, or financing. Its measured output is a resource-constrained investigation frontier. The scientific subtitle names the broader question of economic opportunity; the demonstrated quantity is not a market valuation.

1.2 Contributions and claim limits

The deliverable contains a reproducible capability-construction mechanism, emitted programs, a complete proposal history, fixed evaluation partitions, causal interventions, a second construction, and an execution record from released project code. It also contains the full failure evidence.

The distinctive contribution is the combination of those elements around a portfolio-level frontier test. Principal-component reconstruction, regularization, greedy experimental design, affine calibration, exponential fitting, and program emission are established methods. We do not claim a new sensing theorem, an exhaustive priority claim, a general scientific inventor, or a production-admitted autonomous enterprise.

2 Scientific foundations

Autonomous scientific experimentation has substantial prior art. Robot Scientist Adam combined hypothesis generation, experimental execution, and formalized records (King et al., 2009). A mobile robotic chemist demonstrated autonomous experimental search using laboratory equipment (Burger et al., 2020). Coscientist combined language-model planning, documentation, code, and experimental automation (Boiko et al., 2023). These systems establish that automated scientific loops are not new. Our focal variable is the reusable measurement capability selected before a portfolio is evaluated.

Program-synthesis and library-learning research already studies reusable computational structure. DreamCoder learns libraries and search mechanisms across tasks (Ellis et al., 2021); Stitch synthesizes abstractions from supplied primitives (Bowers et al., 2023). That distinction matters here: autonomous composition inside a declared grammar does not imply that the grammar, physics, or fundamental algorithms were invented autonomously.

Measurement choice can be expressed through explicit objectives, as in experimental-design research (Chaloner and Verdinelli, 1995). Reduced measurements can also support signal reconstruction under suitable structural conditions (Candès et al., 2006). The present decoder uses an empirical low-rank model and ridge inversion. It is not the sparse-recovery procedure in that theorem, and its performance does not inherit the theorem’s guarantee.

Physical sensing further depends on illumination, throughput, saturation, and signal-dependent noise. Multiplexing advantages can change or disappear under those constraints (Ratner et al., 2007). We therefore use a transparent digital exposure model, bound measurement-row norms, charge complementary signed readings separately, and include a signal-dependent-noise stress condition. Those precautions delimit the simulation; they do not establish an optical advantage.

3 A measurable opportunity frontier

3.1 Tasks, capabilities, and resource limits

Let a fixed portfolio be $\mathcal{T} = \{T_i\}_{i=1}^N$. Task T_i specifies an input distribution, a required output, a quality predicate Q_i , and a resource allowance b_i . A capability c is an executable acquisition-and-reconstruction program together with its fitted parameters and a source identity. It does not include a transferable permission to execute a new consequential action.

The first campaign contains $N = 2,048$ missions. Three static families contain 512 missions each; a withheld paired-change family contains another 512. A static mission requires one sample session; a paired mission requires two. With $v_i \in \{1, 2\}$,

$$S = \sum_{i=1}^N v_i = 2,560.$$

Single-session tasks have a 28-exposure allowance; paired tasks have a 56-exposure allowance. Those limits and the quality requirements are fixed before scored evaluation.

Let $D(c)$ be the complete counted discovery charge and $u(c)$ the operating exposures per sample session. The declared equal-per-session allocation gives

$$K_i(c) = v_i \left(u(c) + \frac{D(c)}{S} \right). \quad (1)$$

The observed affordable frontier is

$$\mathcal{F}(c; \mathcal{T}) = \{i : Q_i(c, T_i) = 1 \wedge K_i(c) \leq b_i\}. \quad (2)$$

This is a finite observed set. A failed attempt is not a theorem that the task is impossible, and membership on one sample is not a distribution-wide quality guarantee.

3.2 Conservation of the shared charge

The allocation has a simple audit property:

$$\sum_i K_i(c) = D(c) + S u(c). \quad (3)$$

Every failed proposal, reference acquisition, and current verification charged to $D(c)$ appears once in the campaign total. Paired missions receive twice the fixed-cost allocation because they use the instrument twice. No hypothetical future customer base is used: all 2,560 sample sessions are actually evaluated.

We also report the operating-only frontier, which asks what is affordable after construction is treated as sunk. Keeping it separate prevents a cheap operating number from concealing an expensive discovery process.

3.3 Break-even conditions

Against an initial instrument with cost u_0 , the constructed instrument reduces campaign exposure use when

$$D(c) + S u(c) < S u_0.$$

For the observed $D = 15,504$, $u = 20$, and $u_0 = 68$, resource parity occurs at 323 sample reuses. Meeting the stricter 28-exposure fully loaded per-session allowance requires at least 1,938 sample

reuses under the declared allocation. These are exposure-accounting thresholds, not predictions of sales, demand, or dollars.

At small reuse counts the fixed cost dominates. A framework that reports only mature steady-state use would miss precisely the construction burden this research is meant to test.

4 The digital discovery laboratory

4.1 Signal and measurement model

The laboratory represents a spectrum $x \in \mathbb{R}^{64}$ on a normalized wavelength grid. Development spectra contain a nonnegative baseline and three smooth Gaussian components. Their centers, amplitudes, and widths vary within the frozen ranges in the executable generator. The system must reconstruct the full spectrum, not merely return one scalar.

For a nonnegative measurement row a , a specimen exposure has the form

$$y = g a^\top x + b + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \quad (4)$$

where $g \sim U[0.85, 1.15]$, $b \sim U[-0.04, 0.04]$, and $\sigma = 0.006$. Gain and offset are fixed within a sample session and independently redrawn for another session. A two-exposure signed row measures its nonnegative positive and negative parts separately and subtracts them. Both noise draws and both exposures are retained.

Rows satisfy $\|a\|_2 = 1$. This prevents free algebraic amplification. It does *not* equalize photon dose, optical throughput, or illumination time. In particular, equal L_2 norms need not imply equal L_1 throughput. An exposure is a unit of this digital detector model.

Calibration is also constructed. The grammar allows no calibration, two dark readings, or affine calibration using two dark and two known-standard readings. Standards are developer-supplied measurement primitives. The constructor selects their use; it does not discover metrological traceability. The affine mode costs four exposures in every new session. Calibration observations are not inherited from unrelated sessions.

4.2 Reference acquisition and learned reconstruction

The constructor acquires 24 reference spectra through the same noisy affine full scanner. Each reference costs $64 + 4 = 68$ exposures, for a total of 1,632. The constructor receives those measured reconstructions rather than simulator truth. Ground truth remains available to the synthetic evaluator for scoring.

The mean μ and principal basis B are fitted to the noisy reference library. For an m -row measurement matrix A , the associated rank is $r = \min(m, 16)$. This rank is selected indirectly through the chosen measurement count; it is not a separately invented representation dimension. The emitted decoder is

$$\hat{x} = \mu + B \left((AB)^\top AB + \lambda I \right)^{-1} (AB)^\top (\tilde{y} - A\mu), \quad \lambda = 10^{-4}, \quad (5)$$

where $\tilde{y}$ is the corrected observation vector.

The empirical reference library places 99% of its centered variance in ten principal components. This is a diagnostic about the sampled library, not a guarantee for every future spectrum. Narrow-line stress cases deliberately violate that useful representation.

Dimension	Supplied alternatives
Measurement design	Uniform point samples; data-dependent greedy point samples; empirical principal patterns; cosine patterns; seeded random signed patterns.
Observation count	$m \in \{4, 8, 12, 16\}$.
Calibration	None; two dark readings; four-reading affine calibration.
Reconstruction	The learned mean/basis and ridge inverse in Equation 5.
Budget	25,000 discovery exposures, including reference acquisition, failed proposals, initial diagnosis, and fresh verification.

Table 1: The first construction space contains $5 \times 4 \times 3 = 60$ candidates. Developers supply the primitives and grammar. The winning combination and its fitted numerical implementation are outputs.

4.3 The disclosed construction grammar

Greedy point selection adds the row that maximizes a regularized log-determinant criterion on the fitted basis. Principal patterns use learned basis vectors. Cosine and random patterns provide different projection alternatives. Every candidate is serialized, evaluated on the same permitted validation portfolio, and retained in the history with its outcome and cost.

The selected instrument is **greedy_points-16-affine**. Its 16 point observations plus four calibration readings cost 20 exposures per session. The chosen zero-based channel indices, in construction order, are:

30, 33, 15, 11, 49, 27, 19, 46, 37, 7, 42, 23, 53, 4, 61, 13.

Those indices, the measured mean, and the decoder matrices are emitted in an executable Python module. That emitted module is used for the primary evaluation and for successor construction, rather than being included only as a decorative export.

5 What is autonomous, and what is supplied

The strongest supported autonomy claim is *bounded executable capability composition*. After the source/protocol freeze, the program acquires references, diagnoses the declared resource conflict, constructs alternatives, selects an instrument from validation evidence, emits it, executes it on new requests, and uses it in a second construction without a developer selecting the winning candidate.

The developer supplies the scientific setting, distributions, primitive algorithms, grammar, budgets, standards, scientific objectives, and evaluators. The system does not invent the meaning of spectroscopy, propose a new physical law, or decide which human goals are valuable. It discovers a design choice and fitted implementation within those boundaries.

The six Out-stages organize actual events in the laboratory. Out-Learn acquires the measured reference library. Out-Think records the common cost constraint and empirical structure. Out-Design constructs and scores the candidates. Out-Strategise chooses a design under the fixed requirements and allocates shared costs. Out-Execute invokes the emitted instrument on fresh requests. Out-Evolve constructs and evaluates a different temporal capability. The Python program emits these events as they occur; the workbench records those events rather than manufacturing a stage narrative afterward.

6 Protocol and capable comparisons

Object	Supplied by developers	Produced by execution
Mission portfolio	Required spectra, derived quantities, budgets and quality thresholds	Fresh mission outcomes and affordability membership
Instrument family	Five measurement designs, four counts, three calibration options	Selected design/count/calibration; measured mean, basis and decoder
Search policy	Exhaustive policy and comparator policies	Proposal history, failures, stopping condition, chosen artifact
Successor	Temporal generator, marker, candidate schedules and estimators	Selected times and estimator; emitted temporal program
Lineage	Artifact formats and checks	Parent/child content hashes and current task evidence
Economic meaning	Commercial target and accounting categories	No verified proceeds/offers; USD and human costs remain unknown

Table 2: The autonomy boundary is part of the result. A disclosed construction space is compatible with autonomous selection; it is not open-ended invention.

6.1 Frozen requirements and procedural separation

The protocol and all five laboratory source files were frozen at 19:48:32 UTC on 23 October 2026, before the first scored campaign. The protocol SHA-256 is:

312cff24f05a8ab5744850e83cc611e290c079a75ea01144b41446797aa92162

The freeze is an internal content commitment, not external preregistration. Developers know the generator and seeds. Builder functions receive the permitted references and validation cases; code inspection verifies that held-out cases are generated and evaluated only after the selected plan and emitted program are frozen. This is procedural separation rather than a security sandbox or externally blinded trial.

The selection portfolio contains four cases for each static family, twelve total. A candidate qualifies with at least eleven passes and at least three in every family. Among qualifying candidates, the policy minimizes operating exposure cost, then validation reconstruction error, then canonical candidate identifier. If none qualifies, the program emits a clearly unqualified fallback. Twelve fresh static checks are then charged separately. All three selected builders pass those twelve checks.

Every static mission requires a full reconstructed spectrum with relative L_2 error at most 0.08, plus a family-specific requirement:

Family	Scientific quantity	Error threshold
Total abundance	Sum of reconstructed intensities	Relative error ≤ 0.06
Band balance	Fraction of intensity below grid coordinate 0.5	Absolute error ≤ 0.06
Dominant-emission centroid	$\sum_j \lambda_j x_j^4 / \sum_j x_j^4$	Absolute error ≤ 0.025
Withheld paired change	Centroid of a newly added smooth component, inferred from two reconstructions	Location error ≤ 0.025 ; change-spectrum relative error ≤ 0.25

Table 3: Requirements are fixed before evaluation. The centroid is not the argmax peak position. The full-spectrum requirement is part of each static deliverable, so the comparison does not characterize optimal scalar-only estimators.

The paired-change family introduces a different inference objective and a new component whose location is not restricted to the three customary centers. It receives no bespoke instrument rebuild. It remains within the same simulated spectroscopy environment; this is not a claim of transfer across unrelated physical domains.

6.2 Builders, not deliberately weak solvers

The original scanner is the initial apparatus, not the only baseline. Two additional builders receive the same measured library, validation access, construction grammar, shared-infrastructure rights, and 25,000-exposure cap.

The **independent agent builder** is a deterministic policy authored by a separate AI agent before results were revealed. It searches candidates by cost, completes the least-cost tier containing a qualifying design, and selects by validation error. This is an actual separately authored builder exercise; it is not a general-purpose online LLM-agent benchmark.

The **expert-derived analytical comparator** orders the same candidate space by cost and a first-order error approximation derived from the reference library, then stops at the first qualified design. It is a software proxy. No human spectroscopy expert participated, and the requested strong human-expert contest therefore remains unperformed.

Both comparators can construct and reuse the same kinds of infrastructure as the foundry. Their selected artifacts happen to be identical to the foundry’s artifact. This coincidence is informative: the main advantage lies in acquiring a reusable instrument, not in a unique search intelligence demonstrated by the exhaustive policy.

6.3 Controls and stress

The apparatus-removal contrast returns to the original scanner and assigns it zero discovery overhead, favoring the control. It does not forbid alternative reconstruction strategies in the capable-builder comparisons.

Two component interventions retain the selected design: remove affine calibration, or circularly mismatch the decoder columns with the unchanged measurement rows. The latter keeps detector costs identical while breaking the learned measurement-to-spectrum correspondence.

Two predefined stress conditions test the model boundary. One adds previously unseen narrow spectral lines. The other changes both specimen and bright-standard noise variance by adding 0.0025 times the relevant signal. This is a synthetic heteroscedastic Gaussian stress, not literal photon-count validation.

7 Results: the frontier expands, the builders tie

7.1 Discovery and complete campaign cost

Of 60 proposals, three satisfy the development qualification rule and 57 do not. The exhaustive foundry selects a 16-point affine-calibrated instrument. Its discovery ledger contains 1,632 reference exposures, 816 initial-diagnostic exposures, 12,816 candidate-validation exposures, and 240 fresh-verification exposures:

$$D(C_1) = 1,632 + 816 + 12,816 + 240 = 15,504.$$

Every rejected candidate remains included. The same final instrument is found after 42 candidate evaluations by the independent builder and after 39 by the analytical comparator.

The foundry’s 66,704 exposures fit the full campaign allowance of $2,560 \times 28 = 71,680$. The original apparatus requires 174,080. The reduction of 107,376 exposures is 61.68% of the original total. This resource saving does not imply a 61.68% dollar saving, equal scientific quality, or a commercial return.

All construction costs remain in the ledger

The same 2,048 missions / 2,560 sample sessions; simulated detector exposures

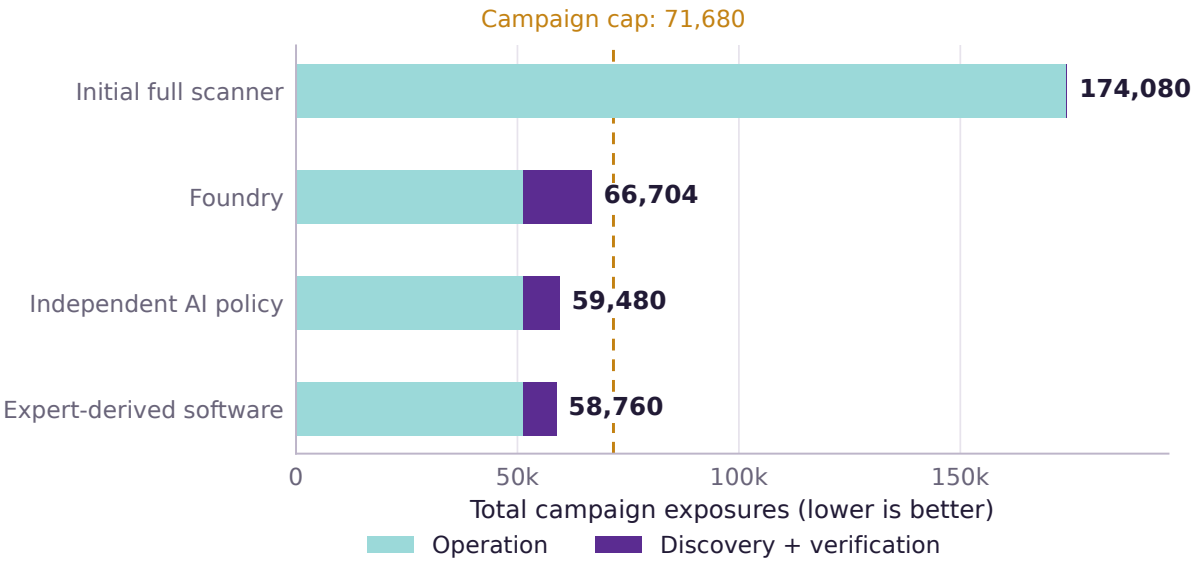
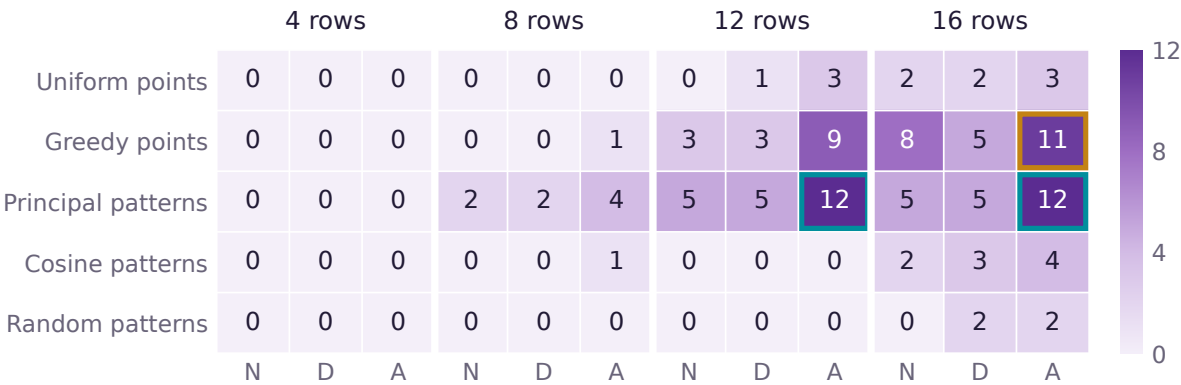


Figure 2: Shared construction must pay for itself. The 51,200 operating exposures are identical for the three constructed-instrument methods. The difference between them is discovery expenditure. No failed proposal or failed downstream mission is removed from the totals.

The full construction space is visible

60 candidates · 57 failed eligibility · 3 eligible · one selected



Cell value: validation missions passed out of 12. N = none; D = dark; A = affine calibration.
Eligibility also requires $\geq 3/4$ in every family. Teal outline: eligible. Amber: selected.

Figure 3: The candidate history is evidence, not a scripted reveal. All 60 first-generation configurations are scored on the fixed development portfolio. Qualification requires both the aggregate and per-family rules. The executable records retain the per-case scores, unsuccessful candidates, costs, and canonical tie-breaking decision.

Method	Discovery	Per session	Campaign	Feasible / 2,048
Initial full scanner	0	68	174,080	0
Possibility Foundry	15,504	20	66,704	1,678
Independent builder	8,280	20	59,480	1,678
Analytical expert proxy	7,560	20	58,760	1,678

Table 4: All costs are simulated detector exposures. Campaign figures include operating costs for failed as well as successful tasks. The foundry reduces use relative to the initial scanner but loses the search-efficiency comparison to both capable builders.

7.2 Fresh missions and withheld work

Family	Missions	Initial quality	C_1 quality	C_1 feasible
Total abundance	512	512	488	488
Band balance	512	512	494	494
Dominant-emission centroid	512	512	492	492
Withheld paired change	512	296	204	204
Total	2,048	1,832	1,678	1,678

Table 5: Initial quality passes do not meet the declared budgets. Conversely, crossing the cost frontier does not erase the constructed instrument’s quality disadvantage. The two capable builders have exactly the same quality outcomes because their selected instrument bytes match.

On the three static families, 1,474 of 1,536 tasks pass. On the withheld family, only 204 of 512 pass: 39.84%, with a conditional Wilson 95% interval of approximately 35.69–44.14%. The unchanged instrument thus supports a measurable subset of new investigations, not reliable transfer of the entire family.

For the original scanner, 296 of 512 paired-change tasks pass quality, but each pair costs 136 exposures against an allowance of 56. Even the higher-resolution apparatus does not pass every paired-change task because each pair has independent calibration noise and a relatively small added signal.

The intervals describe fresh specimen/noise sampling conditional on this one constructed instrument. There is one discovery-seed realization. The mechanical integration replay is not a second independent experiment, and these intervals do not quantify uncertainty over new discovery runs, future domains, or alternative grammar choices.

7.3 Causal interventions and failures

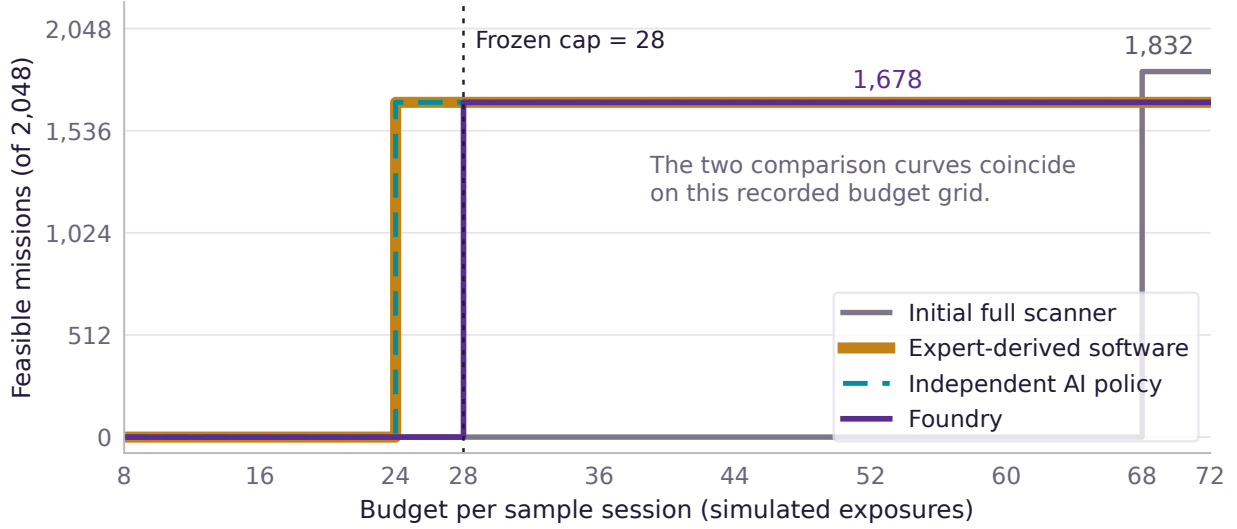
Returning to the initial scanner removes the affordable frontier at the fixed budgets, although it improves quality on many tasks. Removing affine calibration leaves 518 feasible missions. Mismatching the decoder with the measurement rows leaves none. These controls support the importance of the selected measurement/reconstruction/calibration bundle within this simulator.

They do not establish that no alternative instrument could work. Indeed, both capable builders find the same useful bundle. Nor does the fixed original scanner constitute a universal lower bound on measurement cost.

The narrow-line stress is a direct failure: the constructed instrument passes zero of 128 cases while the full scanner passes all 128 quality checks. Under signal-dependent noise, the constructed instrument passes 30 of 128 and the full scanner passes 59 of 128. Those stress figures are quality outcomes; the original scanner remains above the primary task’s resource cap.

Opportunity depends on the budget

Full discovery cost allocated over the frozen 2,560-session campaign



Budgets shown every 2 exposures; connecting steps visualize the recorded grid.

Figure 4: The resource frontier is explicit. Fully counted feasible missions are plotted against the allowed cost per sample session; paired missions receive twice that allowance. The plotted two-unit grid makes the two stronger builders overlap. They reach the same quality-limited plateau at lower allowances. At a large enough allowance, the original scanner supports more quality-qualified missions. Operating-only counts are retained separately in the data.

8 A successor that constructs different work

8.1 From a spectrum interface to a temporal instrument

The second capability estimates a temporal decay rate. A controlled trace has

$$x(t) = x_{\text{background}} + \alpha e^{-kt} m, \quad (6)$$

where m is a supplied smooth spectral marker and the static projected background is unknown to the estimator. The parent instrument supplies reconstructed spectra at proposed times. A fixed marker projection converts those spectra into temporal readouts.

The developer supplies the generator, marker, rate range, and estimator primitives. The system constructs a new executable composition by choosing times and an estimator from observations. This is different work from static spectral reconstruction; it is still bounded composition in a specified scientific family.

Construction first acquires six controlled traces at seven allowed times:

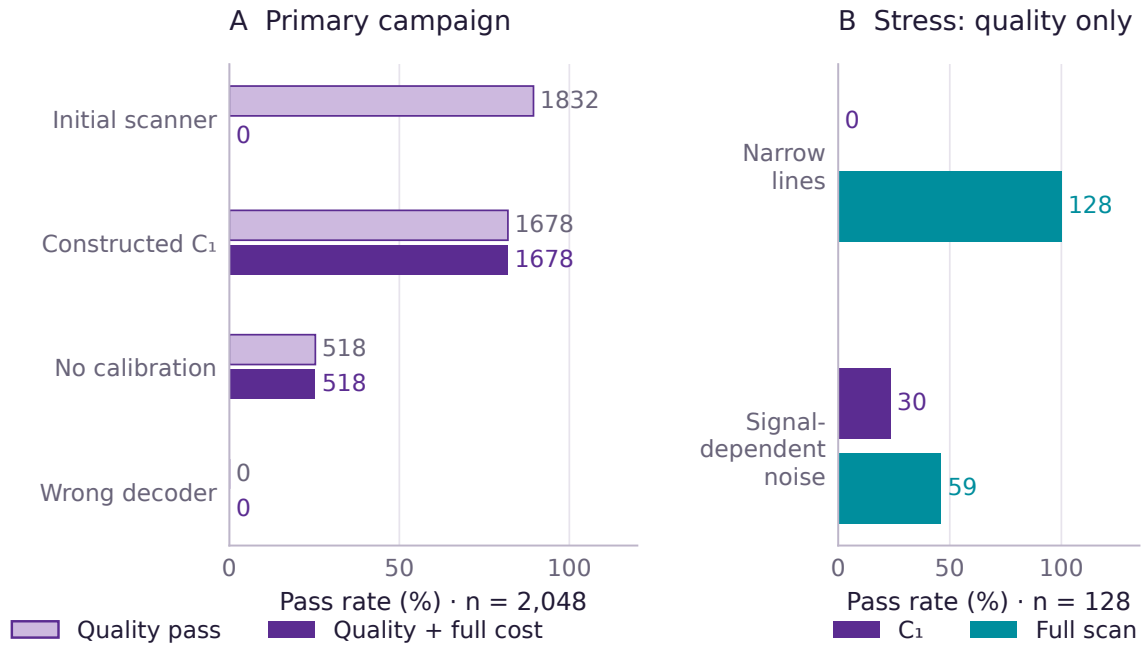
$$0, 0.5, 1, 2, 3.5, 5, 8.$$

The actual emitted parent module performs these 42 acquisitions. At 20 exposures per snapshot, construction data cost 840 exposures. Six time subsets are combined with two estimator primitives: a zero-background log-linear fit and a grid fit of an exponential with nonnegative amplitude and unknown nonnegative background.

All twelve candidates reuse the same acquired traces, so they add computation but no new detector exposures. Their 41,436 grid-point evaluations and 36 log-linear fits are separately recorded. Known

The gain has a defined failure boundary

Quality and affordability are different tests; stress outcomes stay outside the primary frontier.



Labels count passed cases. Full scanning costs 68 exposures/session; the frozen cap is 28.

Figure 5: The boundary of the capability is visible. The component interventions test affordable missions on the primary portfolio. The separate stress panels report quality pass rates on 128 cases per condition. In particular, a smooth-spectrum capability does not become a general spectrometer merely because it has a lineage record.

controlled training rates are provided as calibration labels. Fresh evaluation rates are never supplied to the estimator.

8.2 Selection, execution, and current evidence

Six of twelve candidates fail the all-six-traces construction rule. The selected program uses times $\{0, 2, 8\}$ and the offset-exponential grid estimator. The emitted temporal program is executed on 128 fresh traces. All 128 meet the fixed 15% relative-rate-error requirement.

The child artifact names the parent program's content hash. Its own source, plan, acquired traces, candidate outcomes, and fresh task results are recorded separately. Historical knowledge and source identity transfer; a parent's evidence is not relabeled as a current child check, and no execution authority is inherited.

8.3 The ancestor cost changes the answer

The successor operates with three parent acquisitions per mission, costing 60 exposures. Its construction overhead is $840/128 = 6.5625$, making the marginal fully loaded cost 66.5625 per task, below the 120-exposure allowance.

If the parent must be built from scratch for this small successor campaign, the correct charge is

$$60 + \frac{15,504 + 840}{128} = 187.6875,$$

which exceeds the same allowance. The marginally affordable count is 128/128; the de-novo count is 0/128. Ancestor costs do not disappear simply because the artifact is called a successor.

For a combined history that already executed the full first campaign, the parent discovery is recorded once: 66,704 first-campaign exposures plus 840 child-construction and 7,680 child-operation exposures gives 75,224. This combined total is a different accounting question from a new 128-task business starting from nothing.

8.4 A causal construction control with a precise limit

The same seven-time-point constructor, using the original full scanner, stops after 1,768 exposures before exceeding the 1,800 construction cap. It has only three complete traces and a partial fourth; the next 68-exposure snapshot would exceed the cap. The record says *construction budget exhausted*, not *no solution exists*.

The intended 42 full-scanner snapshots would cost 2,856 exposures, compared with the parent's 840: a 70.59% reduction in this fixed data-acquisition routine. However, an alternative constructor that directly chose three times for six traces would cost $6 \times 3 \times 68 = 1,224$ and fit the construction cap. Our ablation does not exclude that alternative. The three-snapshot full-scanner operating cost would still be 204 per temporal task, above the 120-exposure allowance.

This distinction prevents a truncated routine from being mistaken for a lower bound. The evidence establishes a useful construction dependency and lower data-acquisition cost in the specified routine, not the necessity of this exact inheritance mechanism for all temporal science.

9 Execution through the existing system

9.1 Actual released-code integration

The new laboratory executes through the unchanged AGI ALPHA v1.39.0 workbench at commit `e7fbd7e7fd0599f66c70ea8fe5e26b178e04f541`. Its actual `createWorkbench`, `run`, and `inspectPack` APIs validate the source-bound mission, invoke the Python laboratory, retain artifacts and execution events, and reopen the exported pack.

The adapter verifies thirteen released source, license, and provenance files against their public Git identities. It snapshots the laboratory and protocol before invocation and verifies those inputs again afterward. The vendored Jobs compiler is the unchanged MIT-licensed implementation originally pinned at `ead296b4f515580c108ebae13e993eb4a83d780d` in AGIJobsv0.

The generic workbench plan retains example-specific aggregation wording. We do not present those generic draft obligations as spectroscopy acceptance. A separate exact-purpose scientific constitution is compiled through the same actual Jobs compiler into 21 zero-cap, unsigned, uncommissioned work orders.

The reopened workbench pack is explicitly an unauthenticated historical import with empty active proof and active authority. Its status remains `HOLD`; settlement remains `NOT_CONNECTED`. These are successful preservation of the intended boundary, not evidence that a four-repository commissioned enterprise has completed deployment.

9.2 An integration failure is retained

The first laboratory run completed, but the workbench rejected structured duration metadata because its canonical JSON contract permits integers rather than fractional numeric values. The correction

Component	Executed here	Boundary retained
AGI ALPHA	Released workbench runs the new adapter and laboratory; records actual six-stage events and artifacts	This research adapter is not a production-admitted instrument
Jobs	Unchanged compiler emits spectroscopy-specific bounded drafts	No live job, escrow, contract call, or settlement
SUCCESSOR lineage	Parent/child artifact identities and fresh numerical evidence; pack inspection	Newer held Harness integration is not imported or merged
Manager	Verified role and source boundary documented	Released finite CSV semantics do not certify spectroscopy
Fleet	Verified role and source boundary documented	No native spectroscopy receiver or installed fleet execution
\$AGIALPHA / USDC	Zero-cap AGIALPHA-denominated draft semantics remain distinct from the separate USDC route	No token utility activated, funds moved, approval, or wallet action

Table 6: Executed integration is stated component by component. A compiled draft and an integrity-checked pack do not supply current scientific acceptance, institutional review, or authority.

changed exactly two wrapper expressions to round millisecond durations. Scientific floating-point arrays remained unchanged inside artifact strings.

The laboratory source and protocol were not changed after results. A second execution replayed the same frozen experiment and successfully created and reopened the workbench pack. The original failed integration, original wrapper snapshot, exact patch, and both laboratory evidence sets remain in the package.

This replay is a representation-repair verification, not an independent scientific replication. The counted experimental exposure total is 480,736 for each full run, including comparison policies, controls, stress, and successor work. Both executions therefore account for 961,472 modeled exposure operations in the delivery history.

10 Economics: resource value is not realized profit

10.1 Three distinct ledgers

The first ledger is the standalone candidate campaign, which asks what one builder would spend to provide the declared work. Each comparison pays for its own reference acquisition in that counterfactual accounting.

The second ledger is the actual whole computational study. Its reference library is acquired once and passed to all builders. It includes all candidate validations, fresh checks, four primary methods, two component ablations, stress conditions, the child construction, the truncated full-scanner construction control, and the child evaluation.

The third ledger is monetary. It retains unknown provider charges, hardware allocation/depreciation, and human labor rather than entering zero. Development by AI agents and user-supplied research direction are disclosed. There is no invoice-level audit that proves total attributable cost below \$100.

One complete study execution	Detector exposures
Shared reference acquisition	1,632
Initial scanner diagnosis	816
All builder candidate validations	24,912
All builder fresh checks	720
Four primary-method evaluation campaigns	327,680
Two component ablations	92,160
All stress evaluations	22,528
Successor construction	840
Full-scanner construction control	1,768
Successor fresh evaluation	7,680
Total	480,736
Two retained executions	961,472

Table 7: Complete exposure accounting for the research, separate from the cost of deploying just one selected instrument. CPU work, provider billing and human effort are different quantities.

10.2 The commercial target is not passed

The predefined target is at most \$100 in total attributable cost producing at least \$10,000 in genuinely realizable incremental value. This package establishes neither a verified executable offer nor a settled payment. Its recorded proceeds and offers are zero, and its complete monetary cost is unknown. The target status is therefore **NOT ESTABLISHED**.

Assigning a speculative price to every successful synthetic mission would not repair the missing evidence. Neither would calling exposure savings profit. The work demonstrates a technical resource opportunity that could motivate a later demand-backed test. It does not establish that the simulated investigations have buyers or that the instrument is physically useful.

The existing instrument can make the marginal successor campaign affordable because a productive asset is already available. That is a legitimate conditional result. A new venture without that asset must include the ancestor construction bill, and fails the stated successor allowance in this experiment.

11 Limitations and falsifiable next evidence

Finite construction space. The system selects and instantiates supplied methods. Broader capability classes, new primitive discovery, and open-ended scientific invention are outside the evidence.

One discovery realization. Fresh mission cases are numerous, but only one reference-library/selection-seed pipeline is evaluated. The replay uses the same seeds. Independent runs with new construction libraries and new validation portfolios are needed before generalizing search reliability.

Partial transfer. The withheld paired-change family has 39.84% success. The narrow-line stress has zero success. A future capability must improve those outcomes on a newly frozen evaluation bank; tuning to these disclosed failures would be exploratory.

Comparison strength. Both software builders are allowed shared infrastructure and beat exhaustive discovery cost. However, no independent human domain expert or unrestricted contemporary tool-using agent has contested the design. A stronger builder could discover a cheaper or more accurate interface beyond this grammar.

Synthetic measurement resources. A detector exposure is not a physical unit of energy, photon dose, elapsed laboratory time, or money. Additive Gaussian noise and the affine session model omit drift, wavelength-dependent gain, actuation costs, detector saturation, spectral throughput, and calibration-traceability burdens.

Specific scientific outputs. Full-spectrum reconstruction is mandatory here. A mission requiring only a scalar could admit a cheaper direct estimator. The frontier cannot be transferred to such a different specification without a fresh comparison.

Small development bank. Twelve validation tasks and twelve fresh checks do not certify population-level reliability. The candidate-history transparency permits scrutiny of selection overfitting, but does not eliminate it.

Construction-control scope. The successor ablation fixes a seven-snapshot data-acquisition routine. It does not compare every possible temporal constructor. An explicit cheaper three-snapshot alternative fits the construction cap, although its full-scanner per-mission operating cost remains too high.

Execution and authority. Released Alpha orchestration and real draft compilation are executed. Native spectroscopy semantics for Manager/Fleet, installed execution, the held newer Harness lane, contract calls, and applicable token use are not established by those executions. Artifact identity is not scientific truth or legal authority.

Commercial validity. Demand, willingness to pay, accepted deliverables, rights to physical designs, realized proceeds, and complete attributable dollars remain unmeasured. The language of an enterprise does not remove those empirical obligations.

The next decisive physical experiment

Freeze a bench instrument's true exposure time, photon/dose constraints, wavelength-dependent noise and calibration process. Give an independent domain expert and a conventional tool-using agent the same construction rights. Acquire new calibration specimens, construct the chosen design, then test untouched scientific requests and an untouched temporal family. Record all human and provider costs and evaluate value only against verified demand. Success or failure on that experiment would carry a meaning this digital study cannot supply.

12 Conclusion

The Possibility Foundry turns a supplied construction space into an executed choice of means. The selected instrument makes 1,678 fixed missions affordable, preserves the cost of every rejected proposal, and supports construction of a different temporal capability. It also loses to capable alternative builders on discovery efficiency, fails many withheld investigations, fails an entire narrow-line stress set, and cannot make its small successor campaign affordable from scratch.

Those boundaries make the result useful. A productive capability is valuable only relative to the work it enables, the resources its construction consumes, the failures it retains, and the evidence each new mission earns. The experimental product is a changed, inspectable set of possibilities.

**Don't just ask what intelligence can do.
Ask what it can build that changes the answer.**

A Reproduction and artifact map

A.1 Run the actual released-workbench integration

The complete source archive contains the frozen protocol, laboratory source, unchanged vendored released code with licenses, emitted programs, both execution evidence sets, independent review records, figures, bibliography, and this LaTeX source.

From the extracted project root, with Python 3 and NumPy available and Node.js installed:

```
python3 -m pip install numpy
node integration/run-workbench.mjs -output /absolute/new/output
```

The output directory must not already exist; its parent must exist. The adapter checks released source identities, invokes the laboratory, stores its logs and evidence, compiles the scientific drafts, and reopens the exported pack. It performs no network laboratory control, wallet operation, payment, or native Fleet admission.

To run the laboratory without the released workbench wrapper:

```
python3 -m lab.run -output /absolute/new/laboratory-output
```

That command executes the same frozen science, but does not itself establish the wrapper-integration claim. Exact recorded scientific bytes used Python 3.12.14 and NumPy 2.3.5; the full second laboratory execution took approximately 7.65 seconds in the provided environment. Timing is descriptive and is not a hardware-normalized speed benchmark.

A.2 Key artifacts

Path	Purpose
protocol.json	Fixed task requirements, grammar, budgets, seeds, accounting and monetary-target status.
freeze.json	Protocol and laboratory-source hashes recorded before scored outcomes.
lab/run.py	Construction, comparison, evaluation, interventions, ledgers and actual stage events.
lab/simulator.py	Public digital measurement primitives and evaluator.
lab/comparators.py	Independently authored cost-first and analytical builders, with documented packaging-only changes.
lab/successor.py	Temporal construction, emitted program and ancestor-inclusive accounting.
evidence/workbench-run-02/ laboratory/results.json	Machine-readable final scientific summary.
evidence/workbench-run-02/ laboratory/mission_results.csv	12,288 per-mission rows across four primary methods and two ablations.
evidence/workbench-run-02/ laboratory/foundry_proposal_history. json	Every first-generation candidate's scores and costs.
evidence/workbench-run-02/ laboratory/artifacts/instrument.py	Executable first capability actually used downstream.

Path	Purpose
evidence/workbench-run-02/ laboratory/successor/successor_ program.py	Executable temporal successor.
evidence/workbench-run-02/ integration-receipt.json	Actual released-workbench result and preserved admission/settlement boundaries.
evidence/workbench-run-01/ integration/amendments/	Original successful laboratory and failed wrapper-pack attempt. Exact duration-representation fix and targeted schema reproduction.
reviews/	Scientific, comparator, source, accounting and integration review evidence.
paper/ and paper/figures/	Compilable manuscript, bibliography and vector/raster scientific figures.

The archive includes machine-readable output inventories with file sizes and SHA-256 hashes. Source identity is checked separately from scientific outcomes. Historical imports deliberately have empty current proof and authority.

A.3 Rebuild the publication

With pdfLaTeX, BibTeX, and the declared figure files available, run from `paper/`:

```
pdflatex The_Possibility_Foundry.tex
bibtex The_Possibility_Foundry
pdflatex The_Possibility_Foundry.tex
pdflatex The_Possibility_Foundry.tex
```

The figure-generation script reads retained evidence rather than rerunning scientific selection. The source archive is the authoritative complete reproduction package; an isolated TeX file depends on its accompanying bibliography and figures.

B Interventions, review, and research status

Before outcomes, internal review corrected six specification issues: a centroid was renamed accurately, references were acquired through the noisy scanner, the rank-selection rule was made explicit, full-spectrum deliverables were declared, the measurement-row L_2 bound was distinguished from physical photon fairness, and bright calibration received the stress-noise law. A separate comparator review added explicit fresh-check status, whole-study exposure accounting, and execution of the actual emitted parent module in the successor.

No winning candidate, threshold, evaluation seed, or scientific primitive was changed after the first scored result. The sole post-outcome executable correction was the wrapper's duration representation. Both original and corrected wrapper histories are retained. A fresh numerical replay matched the scientific outputs and selected artifact identities.

Work was conducted by AI-assisted research and engineering roles under the user's direction. Separate agents provided a scientific review, an independently authored comparator, and a released-source integration review. These are same-session work roles, not external peer review, a human-expert trial, or organizational independence. Provider charges and actual human effort are not sufficiently measured to assert a complete cost ceiling.

The artifact is a completed *digital research edition*. It is not a claim that the broader four-repository release effort, native installation, physical laboratory, live job settlement, or commercial target has been completed. Original papers, public URLs, release branches, and recorded restricted operations were preserved.

Project and source links

- AGI ALPHA / SUCCESSOR Ω : <https://github.com/MontrealAI/AGI-Alpha-Agent-v0>
- Public project: <https://montrealai.github.io/AGI-Alpha-Agent-v0/>
- Coordination record: <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/issues/4850>
- Jobs: <https://github.com/MontrealAI/AGIJobsv0>
- Manager: <https://github.com/MontrealAI/AGIJobManager>
- Fleet: <https://github.com/MontrealAI/AGI-Fleet-Linux>

References

- Daniil A. Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624:570–578, 2023. doi: 10.1038/s41586-023-06792-0. URL <https://www.nature.com/articles/s41586-023-06792-0>.
- Vincent Boucher. Invent the economy: Transformational intelligence for creating viable markets. Project research paper, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/main/docs/Invent_the_Economy.pdf. Technical and commercial viability framing; project research context.
- Matthew Bowers, Theo X. Olausson, Lionel Wong, Gabriel Grand, Joshua B. Tenenbaum, Kevin Ellis, and Armando Solar-Lezama. Top-down synthesis for library learning. *Proceedings of the ACM on Programming Languages*, 7(POPL):1182–1213, 2023. doi: 10.1145/3571234. URL <https://arxiv.org/abs/2211.16605>.
- Benjamin Burger, Phillip M. Maffettone, Vladimir V. Gusev, Catherine M. Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M. Alston, Buyi Li, Rob Clowes, Nicola Rankin, Brandon Harris, Reiner Sebastian Sprick, and Andrew I. Cooper. A mobile robotic chemist. *Nature*, 583:237–241, 2020. doi: 10.1038/s41586-020-2442-2. URL <https://strathprints.strath.ac.uk/74759/>.
- Emmanuel J. Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. doi: 10.1002/cpa.20124. URL <https://arxiv.org/abs/math/0503066>.
- Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995. doi: 10.1214/ss/1177009939. URL <https://www.stat.cmu.edu/tr/tr599/tr599.html>.
- Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sablé-Meyer, Lucas Morales, Luke Hewitt, Luc Cary, Armando Solar-Lezama, and Joshua B. Tenenbaum. DreamCoder: Bootstrapping inductive program synthesis with wake-sleep library learning. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pages 835–850. ACM, 2021. doi: 10.1145/3453483.3454080. URL <https://dspace.mit.edu/handle/1721.1/145949>.

Ross D. King, Jem Rowland, Stephen G. Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N. Soldatova, Andrew Sparkes, Kenneth E. Whelan, and Amanda Clare. The automation of science. *Science*, 324(5923):85–89, 2009. doi: 10.1126/science.1165620. URL <https://research.aber.ac.uk/en/publications/the-automation-of-science/>.

MONTREAL.AI. Transformational autonomous enterprises. Writer 4 research edition, 23 October 2026, 2026. URL <https://montrealai.github.io/AGI-Alpha-Agent-v0/papers/transformational-autonomous-enterprises-v1.pdf>. Project research edition; companion execution and publication are distinct claims.

Netanel Ratner, Yoav Y. Schechner, and Felix Goldberg. Optimal multiplexed sensing: bounds, conditions and a graph theory link. *Optics Express*, 15(25):17072–17092, 2007. doi: 10.1364/OE.15.017072. URL <https://doi.org/10.1364/OE.15.017072>.