MONTREAL.AI · QUEBEC.AI

SUCCESSOR Ω × AGI JOBS

The Constitutional Compiler and Proof-Carrying Work System for Mission-Sovereign Intelligence Institutions

From sealed work contracts and sovereign training environments to Specialist ASI, accountable admission and recursive succession

The governing declaration

AGI Jobs manufacture candidacy. The Sovereign Mission Gym manufactures experience. The Sovereign Alpha Foundry competes complete institutions. Fresh comparative proof may establish Specialist ASI. Only accountable admission creates Successor Ω. Chronicle preserves what survives reality and prepares the successor to the successor.

Vincent Boucher

President, MONTREAL.AI & QUEBEC.AI

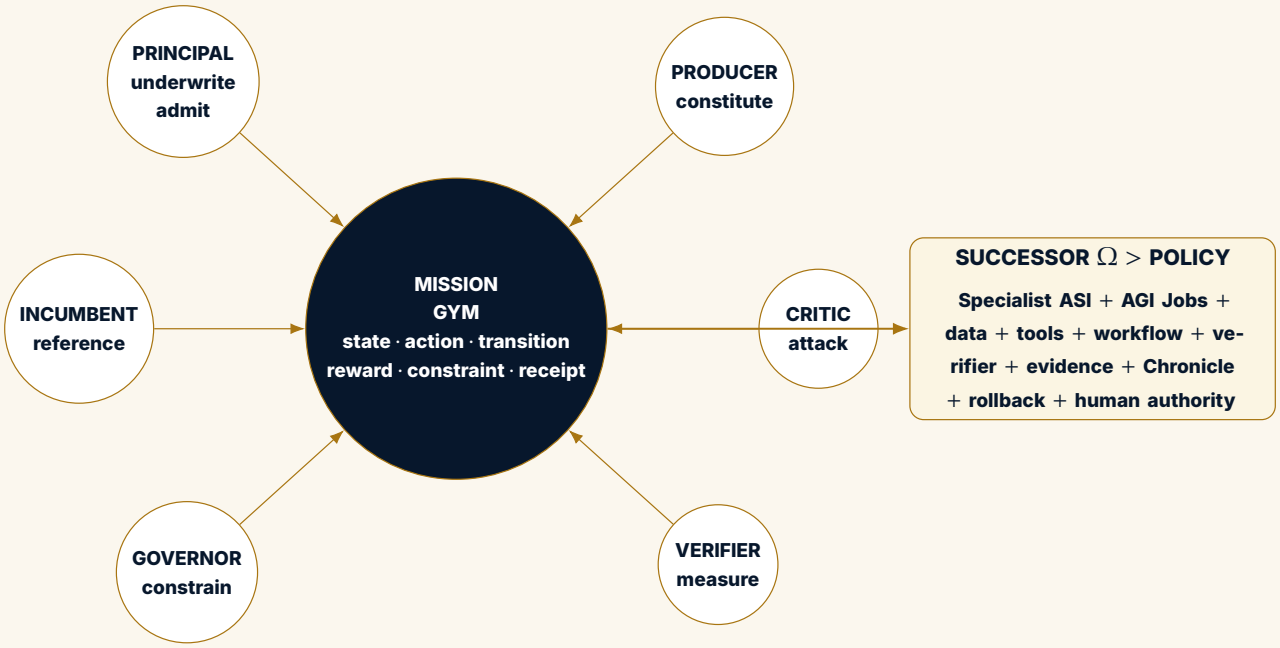
https ://montreal.ai/ · president@montreal.ai

Version 2.0.0 · Dyadic Constitution, Job-Graph Algebra and Synthetic-Evidence Edition · 12 August 2026

Montreal, Quebec, Canada

THE CONSTITUTIONAL OVERTURE

From rented intelligence to proof-bearing institutional agency



Mission Alpha

$$\alpha_m = U(S_m) - \max\{U(I_m), U(B_m)\} - C - R - D - \varepsilon_{\text{gym}}$$

The Spread is an opportunity. Alpha exists only after the complete denominator and protected comparison.

$$A_t \preceq \phi(E_t).$$

Capability may justify candidacy. Fresh proof may establish mission dominance. Only accountable admission may grant authority.

The complete doctrine

Beta is rented. The Spread is underwritten. AGI Jobs constitute the work. The Gym trains and tests. The Foundry competes institutions. Specialist ASI is proven. Successor Ω captures, governs and compounds Mission Alpha while authority remains collateralized by fresh evidence.

The central doctrine

A Successor Ω is not a model. It is the complete mission-bearing institution that has earned comparative superiority and has been separately admitted under a scoped, versioned, monitored, expiring and revocable Authority Envelope. AGI Jobs are the sealed work-and-proof contracts through which that institution is constituted, tested, repaired, requalified and eventually replaced.

Abstract

This paper develops Successor Ω and **AGI Jobs** as the constitutional compiler and proof-carrying work system of the Verified Succession Institution. Its central claim is compositional : a mission-specific intelligence institution is not produced by one prompt, one agent, one model-training run or one benchmark result. It is constituted through a typed Job Graph of bounded work-and-proof contracts that specify objective, principal, identity, authorized inputs, permitted tools, prohibited actions, budget, evidence, verifier, acceptance criteria, rollback, information flow and Chronicle eligibility. The resulting complete architecture is frozen, compared against the incumbent and the best credible alternative on fresh protected work, and considered for a separate accountable admission decision.

The paper formalizes Successor Ω as an admitted institution rather than a capability label. It distinguishes AI Beta, the broadly rentable frontier baseline ; the Mission Advantage Spread, the gross opportunity above the strongest reference architecture ; Mission Alpha, the conservative residual surplus remaining after complete cost, risk, human burden, dependency, Proof Debt and Gym-to-reality basis risk ; Specialist ASI, an earned mission-bounded comparative proof state ; and Successor Ω, the governed institution that captures, constrains and recursively renews that Alpha. The canonical progression is :

Successor Manifold → Mission Advantage Gradient → SEIZE Underwriting → Bounded AGI Jobs → Sovereign Mission Gym → Fresh Proof → Specialist ASI → Successor Ω.

Version 2.0.0 adds a constitutional job compiler ; a work, evidence and authority type system ; proof-carrying work objects ; information-flow non-interference ; authority non-composition ; dependency invalidation ; assurance cut sets ; proof-preserving composition ; a machine-readable contract compiler ; explicit conformance profiles ; SJX-5, a five-mission synthetic architecture market ; and AJGIB-1, a synthetic Job-Graph assurance benchmark. The evidence programmes demonstrate that the rational winner can be deterministic rules, a MuZero/world-model architecture, a sovereign hybrid or no successor at all, depending on mission structure, complete burden and hard gates.

The framework does not assert that universal ASI presently exists, that a real external system has earned Specialist ASI status, or that synthetic results establish production performance. Those states require an exact release, rights-cleared production-representative evidence, independent proof custody, complete-denominator accounting and accountable admission. The governing rule is exact : **authority never exceeds fresh proof.**

Executive summary

Corporations can increasingly rent powerful frontier intelligence, but rentable capability tends to become a common competitive factor. Durable mission advantage requires a different institutional object : a complete successor that controls the mission constitution, work decomposition, evidence, verifier, memory, permissions, portability and operating economics. The transformation is not “adopt a better model.” It is “compile, constitute, prove, admit and continuously renew the best complete architecture for one consequential mission.”

The operating dyad is :

$$\boxed{\text{Successor } \Omega \times \text{ AGI Jobs}} \tag{1}$$

AGI Jobs are sealed manufacturing and proof contracts. Successor Ω is the admitted operating institution. The former defines and verifies the work ; the latter carries the mission, memory and bounded authority. A successor without boun-

ded jobs becomes an opaque capability bundle. Jobs without a successor remain fragmented work without institutional identity, continuity or Alpha capture.

One-line operating doctrine

SEIZE underwrites the intelligence asset. The Constitutional Job Compiler creates the typed manufacturing graph. The Sovereign Mission Gym manufactures experience. Fresh proof may earn Specialist ASI. Accountable admission creates Successor Ω. Chronicle prepares the successor to the successor.

Principal contributions

Contribution	Institutional result
Successor Ω state	Separates capability, proof and admission; defines exact mission identity, evidence, memory, rollback and bounded authority.
AGI Job contract	Converts vague autonomy into typed, auditable work with objective, inputs, tools, prohibitions, evidence, verifier, acceptance, budget, expiry and rollback.
Constitutional Job Compiler	Converts a frozen mission into a typed dependency graph, information-flow policy, proof registry and stop-loss schedule.
Proof-carrying work	Binds output to provenance, tools, actions, interventions, cost, verifier verdict and invalidation dependencies.
Authority non-composition	Prevents job permissions from silently aggregating into successor authority.
Job-Graph assurance	Adds cut sets, verifier-correlation analysis, invalidation propagation and proof-preserving composition.
Sovereign Mission Gym	Supplies replayable, adversarial and measurable experience while keeping certification independent.
Sovereign Alpha Foundry Ω	Routes a portfolio of algorithms and complete architectures by mission structure, not prestige.
Fresh-proof firewall	Freezes one exact challenger and prevents protected information, learning or threshold adaptation from returning to formation.
SJX-5 and AJGIB-1	Provide disclosed synthetic architecture-market and Job-Graph assurance evidence, including negative results and stop verdicts.
Executable institutional twin	Defines machine-readable contracts, manifests, proof receipts, scorecards, admission records and conformance profiles.
Chronicle recursion	Preserves accepted capability and failure so later succession cycles improve without inheriting proof or authority.

Claim boundary

This is a research, institutional-design and reference-implementation paper. “Mission Alpha,” “proof capital,” “Specialist ASI,” “Successor Ω” and related terms are defined operating constructs. Mission Alpha is benchmark-relative, risk-adjusted mission surplus; it is not a securities-return projection. Synthetic cases, algorithm tournaments and simulation gates do not establish real customer value or production authority. A real Specialist ASI claim requires fresh protected mission evidence; a real Successor Ω state additionally requires accountable admission, rights, security, monitoring, expiry, rollback and revocation.

CONTENTS

PART I

Foundations : Mission, Environment and Succession

From static plans to verified succession under capability-time pressure

The institutional operating condition, Navigator Ω, and the mission triad of Gym, Specialist ASI and Successor Ω.

SUCCESSOR MANIFOLD → GRADIENT → SEIZE → BOUNDED JOBS → SPECIALIST ASI → SUCCESSOR Ω

1 The Operating Condition : Capability Time and Institutional Succession

1.1 From stable planning to capability time

A conventional institution assumes that its operating architecture remains valid long enough for annual planning, procurement, implementation and review. That assumption becomes fragile when models, tools, costs, providers, regulations, interfaces and customer expectations change on a timescale shorter than the institution's adaptation cycle. The relevant operating regime is not defined by one metaphysical date. It is defined by a relation between the half-life of accepted assumptions and the lag required to detect, prove and authorize a replacement.

Let H_c denote the expected half-life of a material capability or architecture assumption, and let L_a denote the institution's adaptation lag. Define the **Succession Exposure Ratio**

$$\mathrm{SER} = \frac{L_a}{H_c}. \tag{2}$$

When $\mathrm{SER} < 1$, the institution generally adapts before its assumptions become materially stale. When $\mathrm{SER} > 1$, the institution increasingly acts from depreciated architectures. The objective is not speed alone. An institution can shorten L_a by eliminating proof, rights review, human authority or rollback and thereby accelerate toward a larger failure. The relevant objective is **Verified Adaptation Velocity** :

$$\mathrm{VAV} = \frac{\Delta K_v}{\Delta t} \cdot \frac{1}{1 + \lambda_R R + \lambda_H H + \lambda_D D_P}, \tag{3}$$

where K_v is verified productive capability, R residual risk, H human and reviewer burden, and D_P Proof Debt.

Recent work on possible transitions from AGI to ASI describes scaling, paradigm shifts, recursive improvement and large multi-agent collectives as distinct but potentially interacting pathways, while emphasizing substantial uncertainty and bottlenecks rather than a single guaranteed discontinuity (Genewein et al., 2026). GoalOS therefore treats capability time as an environment to govern rather than a calendar prophecy.

Capability-time law

A capability event does not authorize replacement. It creates a duty to re-underwrite the incumbent, reconstruct the reachable successor set, purchase decisive evidence and preserve the ability to refuse.

1.2 Every incumbent is a depreciating institutional architecture

An incumbent is more than software. It includes the objective, division of labour, data rights, knowledge representation, tools, models, providers, human approvals, verifier, workflow, contracts, infrastructure, evidence, memory, incident routes and authority. Its depreciation may be technical, economic, strategic, legal or constitutional.

Let S_0 be the incumbent. The expected present burden of waiting is

$$C_{\mathrm{wait}}(S_0) = V_{\mathrm{foregone}} + C_{\mathrm{legacy}} + R_{\mathrm{competitive}} + R_{\mathrm{dependency}} + I_{\mathrm{option\ loss}}, \tag{4}$$

while the expected burden of moving too early to candidate S_i is

$$C_{\mathrm{premature}}(S_i) = C_{\mathrm{formation}} + C_{\mathrm{integration}} + R_{\mathrm{error}} + R_{\mathrm{security}} + C_{\mathrm{rollback}} + R_{\mathrm{lock\text{-}in}} + D_P. \tag{5}$$

Verified succession is a comparison between those two burdens, not a choice between enthusiasm and caution.

1.3 Why the Gym becomes necessary

Without an executable mission environment, an institution can compare successors only through rhetoric, isolated demonstrations or historical anecdotes. It cannot reliably answer :

— whether candidates faced the same cases ;

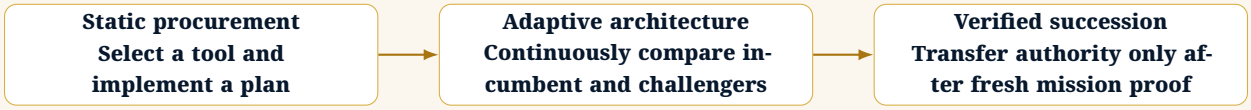
— whether hidden human rescue produced the result;
— whether the score represented the real mission;
— whether rare failures were sampled;
— whether authority and tools were bounded;
— whether the candidate can be replayed after a component changes;
— whether the successor remains superior to a new external alternative.

The Mission Gym converts the Mission Constitution into executable state, observation, action, transition, constraint and verification logic. It makes failure replayable and comparative proof possible. It also becomes a corporate option on every future model or architecture : new components can be substituted and tested against an institution-owned mission rather than accepted on provider reputation.

Design Proposition 1.1 (Gym necessity). Where a mission is valuable, repeated, bounded, measurable and capable of replay or simulation, a sovereign Mission Gym reduces the cost of comparing successor architectures and increases the credibility of authority decisions. The proposition fails when the Gym has greater basis risk than the uncertainty it is meant to resolve.

1.4 The unit of improvement is the mission-bearing institution

A model upgrade can be useful without changing the institution. Conversely, a superior successor may arise from a smaller model, deterministic tool, better verifier, redesigned human role, new contract, different provider, improved data or narrower authority. The unit of succession is therefore the complete mission-bearing institution.



1.5 Claim discipline

A forecast is not a capability result. A Gym score is not a mission result. A mission result is not mission dominance. Mission dominance is not authority. Access is not authority. Historical proof is not fresh proof. A policy is not the complete Successor. These distinctions are the logical foundation of the paper.

No authority by narrative

No forecast, brand, model scale, benchmark reputation, generated explanation, simulation score or strategic urgency can substitute for the exact evidence required by the Mission Constitution. Where evidence is absent, the correct institutional state is uncertainty, not inherited authority.

2 The Successor Ω × AGI Jobs Thesis

2.1 The unit of improvement is the mission-bearing institution

The strongest form of corporate AI is not a larger model placed beside the incumbent workflow. It is a mission-bearing institution capable of receiving observations, performing bounded work, producing evidence, submitting to independent challenge, operating inside a defined Authority Envelope and preserving accepted learning for the next succession cycle. The central unit of improvement is therefore not the policy alone, but the complete architecture that carries the mission.

The Successor Ω × **AGI Jobs** thesis states that every mission-bearing successor must have two mutually constraining structures :

1. an **institutional state** : the mission, identity, architecture, evidence, Chronicle, authority, rollback and accountable principal that define Successor Ω; and

2. a **work constitution** : the portfolio of Bounded AGI Jobs that specify how the institution is formed, trained, tested, operated, repaired, requalified and retired.

A successor without a job constitution risks hidden work, hidden tools, hidden human rescue, evaluator leakage and authority laundering. A job portfolio without a successor risks fragmentation : accepted outputs remain scattered across agents, vendors, prompts and transient conversations, with no durable mission identity, no accumulated evidence and no responsible authority structure.

Definition 2.1 (Successor–Job dyad). For mission m, the Successor–Job dyad is

$$\mathfrak{D}_m = \left(\gg_m^{\Omega}, \mathscr{G}_{J,m}\right), \tag{6}$$

where $\gg_m^{\Omega}$ is the exact admitted mission institution and $\mathscr{G}_{J,m}$ is the typed dependency graph of jobs required to constitute, prove, operate and renew it.

2.2 The three non-equivalent objects

The paper distinguishes three states that ordinary AI programmes often collapse :

A candidate may be technologically impressive yet fail proof. A Specialist ASI may be mission-dominant yet remain recommendation-only. A Successor Ω may operate for one mission and one version while remaining prohibited from acting in another. This separation is the core defense against capability being mistaken for institutional authority.

2.3 The manufacturing identity

Let $\{\mathscr{J}_{m,k}\}_{k=1}^{K}$ be the required AGI Job portfolio and let Assemble integrate the accepted outputs into a complete candidate. Let Prove denote independent fresh comparison, and let Admit denote accountable constitutional admission. Then :

$$\gg_m^{\Omega} = \mathsf{Admit}_m\left[\mathsf{Prove}_m\left(\mathsf{Assemble}_m\{\mathscr{J}_{m,k}\}_{k=1}^{K}\right)\right]. \tag{7}$$

Neither proof nor admission is executed by the candidate alone. The composition is deliberately discontinuous : formation can generate candidacy ; fresh proof can establish mission dominance ; accountable authority alone can create the admitted successor state.

The manufacturing law

AGI Jobs manufacture candidacy. Fresh proof may establish Specialist ASI. Only accountable admission creates Successor Ω.

2.4 Why the multiplication sign matters

The multiplication sign in the title denotes interaction rather than mere addition. AGI Jobs shape the successor, while the successor constitution shapes every job. The mission defines what work may exist ; the job graph defines what the candidate can become ; the verifier constrains what may be accepted ; the Authority Envelope constrains what accepted capability may do ; Chronicle determines which results may influence the next generation.

This bidirectional relation can be written as a coupled system :

$$\mathscr{G}_{J,m,t+1} = \mathscr{C}\left(m, \gg_{m,t}, \mathscr{E}_t, \mathsf{c}_t\right), \tag{8}$$

$$\gg_{m,t+1}^{\text{cand}} = \mathscr{F}\left(\mathscr{G}_{J,m,t+1}, \text{Formation Gym}, \text{resources}\right), \tag{9}$$

$$\gg_{m,t+1}^{\Omega} = \mathscr{A}\left(\gg_{m,t+1}^{\text{cand}}, \text{Fresh Proof}, \text{human admission}\right). \tag{10}$$

The compiler $\mathscr{C}$ is not free to invent authority. The Foundry $\mathscr{F}$ is not free to certify its product. The admission operator $\mathscr{A}$ is not free to ignore proof or rollback. The system remains useful precisely because the powers remain separated.

3 GoalOS Singularity Navigator Omega + SEIZE : The Verified Succession Institution

3.1 The unified institutional object

The merged architecture is not a collection of adjacent tools. It is one recurring constitutional institution whose purpose is to keep consequential missions superior while their capability environment changes. GoalOS Singularity Navigator Ω senses frontier, cost, legal, security and market discontinuities ; the Successor Book maintains the incumbent and the field of possible replacements ; **SEIZE**underwrites one reversible trajectory ; the Mission Gym makes the mission executable and comparable ; the Successor Foundry constitutes materially different complete candidates ; the Independent Proof Plane tests one frozen challenger ; accountable authority admits or rejects ; and Chronicle preserves only the exact capability that survives reality.

Definition 3.1 (Verified Succession Institution). A Verified Succession Institution is a recurring constitutional system that detects when an incumbent architecture has become materially suboptimal, maintains a portfolio of successor options, authorizes bounded proof capital, constitutes complete candidate institutions, compares one frozen challenger on fresh protected work, grants authority only through a separate accountable decision, and continuously re-underwrites the admitted successor.

The institution therefore operates across two coupled loops :

$$\underbrace{\text{Sense} \to \text{Book} \to \text{Underwrite}}_{\text{Navigator } \Omega \text{ macrocycle}} \quad \Longrightarrow \quad \underbrace{\text{Gym} \to \text{Form} \to \text{Prove} \to \text{Admit} \to \text{Requalify}}_{\text{mission succession cycle}} . \tag{11}$$

The macrocycle decides *which mission and which uncertainty deserve institutional attention*. The mission cycle decides *which exact successor, if any, has earned bounded authority*. Conflating these loops creates predictable errors : strategic excitement becomes a development mandate, development becomes deployment, and one attractive output becomes institutional memory.

3.2 Six permanent functions

Function	Controlling question	Canonical output
Sensing	What changed, and which incumbent assumption is impaired ?	Frontier Event Record, Capability Twin update, source and expiry record.
Underwriting	Which successor trajectory is valuable, feasible, governable and worth proof capital ?	Successor Book position, SEIZE memo, proof-capital authorization and stop-loss.
Formation	Which complete architectures should be constituted rather than merely imagined ?	Successor Manifold, Bounded AGI Job portfolio, exact candidate manifests and reserve lineages.
Proof	Does one frozen challenger outperform the incumbent and the strongest credible alternative on fresh work ?	Protected scorecard, proof receipts, lower-confidence result, critical-error record and independent verdict.
Authority	What may this exact release observe, recommend, prepare or execute, for how long and with what rollback ?	Admission Record, Authority Envelope, monitoring, expiry, revocation and known-good fallback.
Compounding	What accepted capability, failure and evidence may shape the next succession cycle ?	Chronicle admission, Negative Capability Graph, Capital-to-Capacity memo and requalification trigger.

These functions are permanent even when the underlying model portfolio changes. A provider may supply intelligence, a consultant may supply expertise, a platform may supply tools and an employee may supply judgment, but none of those roles automatically becomes the institution that decides what counts as truth, authority or reusable capability.

3.3 Navigator Ω as the sensing and portfolio layer

Navigator Ω maintains a live institutional state

$$Z_t = \left(F_t, I_t, \mathscr{B}_t, \mathscr{O}_t, E_t, A_t, K_t\right), \tag{12}$$

where F_t is the external frontier, I_t the accepted incumbent and its assumptions, $\mathscr{B}_t$ the best credible alternative set, $\mathscr{O}_t$ the Successor Book, E_t current evidence, A_t authority and K_t Chronicle-admitted capability. A material event ξ_t updates the state through

$$Z_{t+1} = \mathfrak{N}(Z_t, \xi_t), \tag{13}$$

but the update is not itself a promotion. Navigator Ω can mark an assumption as impaired, open a SEIZE mandate, lower a position, preserve an option or trigger requalification. It cannot certify the candidate or enlarge the Authority Envelope.

A useful sensing layer includes :

— a **Capability Twin** : the institution's current map of models, data, tools, workflows, people, rights, costs, providers, controls and bottlenecks ;
— an **Incumbent and Assumption Ledger** : the claims that justify the present architecture, together with owners, evidence, expiry and impairment triggers ;
— a **Frontier Radar** : current capability, cost, law, security, infrastructure and market events with source quality and decision relevance ;
— the **Successor Book** : incumbent, immediate, challenger, reserve, hedge and retirement positions ;
— an **Opportunity and Risk Brief** : what changed, what matters, what evidence would change the position and what capital or authority is presently exposed.

3.4 SEIZE as the constitutional underwriting microcycle

SEIZE converts one succession event into a bounded proof decision :

S — Surface.
Identify the capability, cost, legal, security, market or infrastructure event ; the assumption that expired ; and the mission newly threatened or made feasible.

E — Evaluate.
Generate materially different retain, repair, rent, configure, form, partner, acquire, hedge, reserve, retire and prohibit alternatives.

I — Instantiate.
Freeze mission, beneficiary, incumbent, best alternatives, hard gates, rights, proof budget, stop-loss, verifier, protected distribution and authority ceiling.

Z — Zero in.
Purchase the smallest evidence object with positive expected decision value.

E — Elevate.
Freeze one exact complete challenger and submit it to independent fresh proof.

Let $a \in \mathscr{A}_{\text{succ}}$ denote a succession action and e a candidate evidence package. The value of proof is

$$\mathrm{VoP}(e) = \mathbb{E}_e\left[\max_a \mathbb{E}[U(a) \mid e]\right] - \max_a \mathbb{E}[U(a)] - C(e). \tag{14}$$

SEIZE buys evidence only when it can change the decision, reduce expected loss, preserve valuable optionality or create transferable institutional capacity. It ends with a Promotion Request, not a Release Certificate.

3.5 The closed-loop succession law

Verified succession law

The frontier creates candidates. Navigator Ω detects and books the opportunity. SEIZE underwrites the smallest reversible proof path. The Gym exposes the mission. The Foundry constitutes complete successors. Protected reality determines mission dominance. Accountable authority admits. Chronicle preserves what survives and opens the next succession event.

The resulting control relation is

$$\boxed{A_t \preceq \phi(E_t) \preceq A_{\max}(m),} \tag{15}$$

where $A_{\max}(m)$ is the mission's constitutional ceiling. Better capability may justify a new proof mandate; it never grants itself additional authority.

4 Successor Ω as the Complete Intelligence Institution

4.1 A system-level definition

Definition 4.1 (Successor Ω). A Successor Ω is an exact, mission-specific, versioned and accountable intelligence institution that has (i) passed fresh comparative proof against the incumbent and the best credible alternative, (ii) been separately admitted by an authorized principal, and (iii) received only the scoped, monitored, expiring and revocable authority supported by current evidence.

A Successor Ω is larger than the policy that selects actions. Let

$$\gg_m^{\Omega} = (M, D, H, T, V, W, R, I, E, C, A, P), \tag{16}$$

where M is the model and policy portfolio; D the rights-cleared data; H the harness and orchestration; T tools, retrieval, deterministic components and planners; V the independent verifier; W the workflow and AGI Job graph; R rights, security, resilience and rollback; I integration; E the evidence record; C Chronicle; A the Authority Envelope; and P the accountable human or institutional principal.

A model can be changed without changing the mission institution if the mission-level observables, evidence chain and authority invariants remain within preregistered tolerance. Conversely, changing a small prompt can create a new release if the change alters action selection, evidence, cost, rights or failure modes. Exact release identity is therefore institutional rather than cosmetic.

4.2 The Omega state

The symbol Ω denotes constitutional closure for one mission and one release, not universal completion. Define

$$\begin{aligned}\Omega_m(x) = \mathbb{1}[\text{MissionDominant}]\mathbb{1}[\text{FreshProof}]\mathbb{1}[\text{HumanAdmission}] \\ \cdot \mathbb{1}[\text{AuthorityIssued}]\mathbb{1}[\text{RollbackReady}]\mathbb{1}[\text{ChronicleBound}].\end{aligned} \tag{17}$$

Only when $\Omega_m(x) = 1$ may the release be called Successor Ω, and only for the mission, scope, duration, jurisdiction, data classes and consequence levels actually admitted.

4.3 Mission sovereignty

Mission sovereignty is not legal personhood, self-ownership or unlimited autonomy. It is the institution's ability to preserve mission continuity and bargaining power while substituting models, providers and components. A mission-sovereign architecture owns or controls the mission constitution, job graph, evidence, Chronicle, permissions, portability and economic capture even when some intelligence components are rented.

The sovereignty vector is

$$\mathbf{s}_m = (s_{\text{mission}}, s_{\text{architecture}}, s_{\text{data}}, s_{\text{evidence}}, s_{\text{memory}}, s_{\text{authority}}, s_{\text{portability}}, s_{\text{economics}}). \tag{18}$$

No one coordinate is sufficient. A fully local model can remain non-sovereign if its evidence, verifier, workflow and data rights are controlled elsewhere. A hosted model can participate in a sovereign successor when it is replaceable and the institution retains the mission-level state.

4.4 Authority as a separate state variable

Capability and authority are separate. Let E_t be current evidence and A_t the Authority Envelope. The constitutional constraint is

$$A_t \preceq \phi(E_t), \tag{19}$$

where $\preceq$ orders consequence, scope, duration, reversibility, transaction size and external effect. A candidate may improve in capability while authority remains unchanged. Historical success cannot justify authority after the release, environment, verifier or evidence state materially changes.

Successor Ω law

Specialist ASI is the proven capability state. Successor Ω is the governed institutional state that carries the mission, evidence, memory, rollback and bounded authority.

5 AGI Jobs as Sealed Work-and-Proof Contracts

5.1 Why ordinary tasks are insufficient

A conventional task requests an output. A Bounded AGI Job specifies the complete conditions under which the work may be attempted, evaluated, accepted, reversed and reused. This distinction matters because consequential machine work involves hidden actions, tool calls, external dependencies, data rights, retry costs, human rescue, evidence quality and authority boundaries that are invisible in an ordinary prompt.

Definition 5.1 (Bounded AGI Job). For mission m, job k is the typed contract

$$\mathcal{J}_{m,k} = (o_k, p_k, i_k, \mathcal{D}_k, \mathcal{T}_k, \mathcal{A}_k^+, \mathcal{A}_k^-, b_k, \tau_k, y_k, e_k, v_k, q_k, \rho_k, \chi_k), \tag{20}$$

where o_k is the objective; p_k the accountable principal; i_k acting identity; $\mathcal{D}_k$ authorized data; $\mathcal{T}_k$ tools; $\mathcal{A}_k^+$ permitted actions; $\mathcal{A}_k^-$ prohibited actions; b_k resource budget; τ_k duration or expiry; y_k output schema; e_k evidence obligation; v_k verifier; q_k acceptance predicate; ρ_k rollback route; and χ_k Chronicle eligibility.

The job is sealed when its principal fields are frozen before execution and its resulting proof receipt binds the exact work, inputs, tools, retries, human interventions, outputs, costs and verifier result.

5.2 Job states

A canonical AGI Job state machine is :

Execution is not acceptance. Acceptance is not Chronicle admission. Chronicle eligibility is not authority. Each transition requires its own evidence and accountable actor.

5.3 Job types

A complete successor requires several distinct job families :

Job family	Institutional purpose
Constitution jobs	Freeze mission, beneficiary, incumbent, alternatives, constraints, rights, proof budget and authority ceiling.
Evidence jobs	Acquire, reconcile and provenance-bind mission facts, examples, failures, outcomes and protected cases.
Formation jobs	Build models, policies, world models, tools, retrieval, workflows, specialists, simulators and integrations.
Challenge jobs	Generate anti-theses, adversarial cases, reward attacks, verifier attacks, failure clusters and stop conditions.
Verification jobs	Apply independent scorers, calibrate evidence, count complete burden, test transfer, rollback and degraded modes.
Admission jobs	Produce accountable acceptance, Authority Envelopes, release certificates, Chronicle decisions and revocation conditions.
Renewal jobs	Monitor drift, impair stale releases, re-underwrite the mission, form descendants and run the next fresh proof cycle.

5.4 The proof receipt

Every job emits a proof receipt :

$$\mathscr{R}(\mathscr{J}_{m,k}) = \text{Hash}(\text{job version, inputs, tools, actions, human interventions, cost, evidence, verdict, rollback}). \tag{21}$$

The receipt turns work into attributable institutional evidence. It also exposes hidden retries, external calls and human rescue that could otherwise manufacture false efficiency.

5.5 AGI does not mean one universal worker

An AGI Job may be executed by a frontier model, compact specialist model, deterministic program, agent team, human expert, vendor, laboratory or mixed organization. The name refers to the constitutional unit of machine-era work, not a requirement that one general intelligence perform everything. A job manager or settlement rail may coordinate assignment and payment, but it does not decide mission truth, acceptance, Chronicle admission or authority.

AGI Job law

An AGI Job is bounded mission work with explicit rules, evidence, verification, acceptance criteria and rollback. It prevents “use AI” from becoming an uncontrolled black box.

6 The Job Graph and Compositional Successor Manufacture

6.1 From a list of tasks to a constitutional dependency graph

The complete job portfolio is not a flat checklist. It is a directed, typed dependency graph

$$\mathscr{G}_{J,m} = (V_J, E_J, \lambda, \kappa), \tag{22}$$

where vertices are AGI Jobs, edges encode evidence and dependency requirements, λ assigns job types and principals, and κ assigns criticality, budget, authority and proof obligations.

A job may begin only when its predecessor evidence is valid, rights-cleared and current. A downstream acceptance may be invalidated when a predecessor is impaired. The graph therefore functions as both manufacturing plan and control surface.

6.2 A canonical manufacturing graph

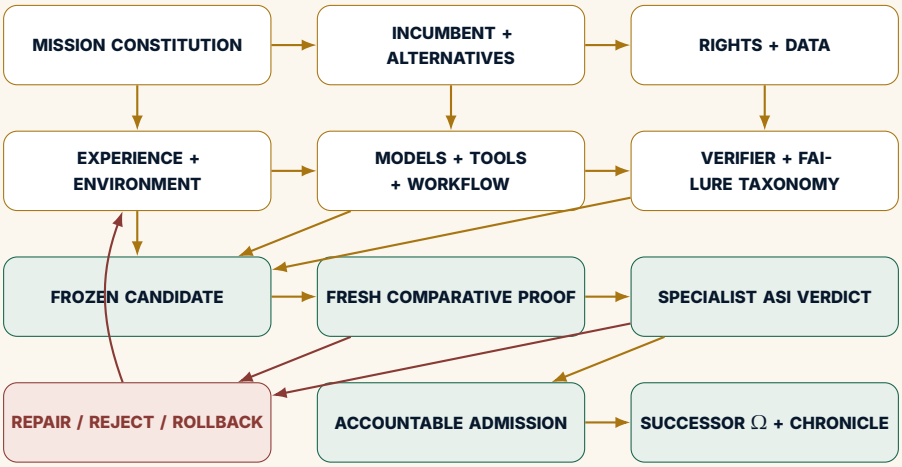


FIGURE 1. The Successor–Job manufacturing graph. Formation creates candidacy; protected proof creates a verdict; accountable admission creates the Successor Ω state.

6.3 Completeness

Let $\mathscr{R}_m$ be the set of mission-critical institutional functions : model, data, orchestration, tools, verifier, workflow, rights, integration, evidence, Chronicle, rollback and authority. Define coverage

$$\text{Coverage}(\mathscr{G}_{J,m}) = \frac{\sum_{r\in\mathscr{R}_m} w_r \mathbb{1}[\exists\, \mathscr{J}_k \text{ accepted for } r]}{\sum_{r\in\mathscr{R}_m} w_r}. \tag{23}$$

Candidate completeness requires full coverage for every function marked non-negotiable by the Mission Constitution. High model performance cannot compensate for an absent verifier, missing rights, untested rollback or undefined human authority.

6.4 Compositionality without authority leakage

Job permissions do not automatically aggregate into successor authority. Let A_k be the action class permitted to job k during formation. The authority of the complete successor is not

$$A^{\Omega} = \bigvee_k A_k.$$

Instead,

$$A^{\Omega}_{m,t} \preceq \phi\left(E^{\text{fresh}}_{m,t}, \text{mission ceiling}, \text{admission record}\right). \tag{24}$$

A job may use a tool inside a sealed sandbox without making that tool available in production. A verifier may inspect protected data without disclosing it to the producer. A formation worker may generate a descendant without holding release credentials. This non-compositional authority rule prevents the job graph from becoming a covert privilege escalation path.

6.5 Critical path and proof capital

The job graph creates a critical path through evidence and formation. Proof capital should be allocated to the job or dependency whose resolution most changes the succession decision. For diagnostic job k, define

$$\text{EVI}_k = \mathbb{E}\left[\max_{a\in\mathscr{A}} U(a \mid e_k)\right] - \max_{a\in\mathscr{A}} \mathbb{E}[U(a)] - C(\mathscr{J}_k). \tag{25}$$

The next job should maximize positive expected value of information subject to rights, safety, time and stop-loss constraints. This makes the job graph a real-options instrument rather than an unlimited research queue.

Compositionality law

Accepted AGI Jobs may compose capability and evidence. They do not compose authority. The complete successor receives authority only through a separate admission record tied to current fresh proof.

7 The AGI Job Contract Calculus

7.1 Why a calculus is required

A job catalogue describes what work exists. A contract calculus specifies when work is admissible, how jobs compose, which evidence is conserved, where failure propagates and why the resulting capability still cannot authorize itself. The relevant comparison is not with ordinary project management. It is with typed interfaces, assume–guarantee contracts, proof-carrying execution and capability-secure systems : each job receives only the information, tools and permissions required by its role, and every accepted transition carries machine-verifiable evidence (Hoare, 1969; Lamport, 1994; Benveniste et al., 2018; Miller et al., 2003).

7.2 Typed signature and operational semantics

For mission m, a job is represented by the typed morphism

$$\mathcal{J}_k : (X_k, E_k, A_k, B_k) \longrightarrow (Y_k, R_k, F_k), \tag{26}$$

where X_k is the authorized input type; E_k the admissible evidence state; A_k the action capability; B_k the resource and time budget; Y_k the output type; R_k the proof receipt; and F_k the declared failure state. A job transition is

$$(x, e, a, b) \overset{\mathcal{J}_k}{\Longrightarrow} (y, r, f) \tag{27}$$

only when the precondition $P_k(x, e, a, b)$ holds. Acceptance requires the guarantee

$$Q_k(y, r) \wedge \mathrm{Verify}_{v_k}(y, r) = 1 \wedge f = \varnothing. \tag{28}$$

The job runtime must fail closed when a required input, right, verifier, budget, credential or source has expired. A plausible output without a valid receipt is an unaccepted output.

Definition 7.1 (Job refinement). Job $\mathcal{J}_j$ refines $\mathcal{J}_i$, written $\mathcal{J}_j \sqsubseteq \mathcal{J}_i$, when it weakens no hard guarantee, expands no prohibited action, preserves the required evidence interface and performs the same mission function with equal or lower complete burden.

Refinement is deliberately stricter than “better benchmark score.” A candidate job that is faster but loses provenance, rollback or independent verification is not a constitutional refinement.

7.3 Assume–guarantee composition

Suppose job i guarantees G_i under assumptions H_i, and job j guarantees G_j under assumptions H_j. Sequential composition is admissible only when

$$G_i \vDash H_j \tag{29}$$

and the output evidence type of i satisfies the input evidence type of j. The composed contract is

$$\mathcal{J}_j \circ \mathcal{J}_i = (H_i,\ G_i \wedge G_j,\ R_i \| R_j), \tag{30}$$

where $R_i \| R_j$ is an append-only receipt chain. This prevents a downstream job from silently assuming facts, rights or authority that no predecessor actually established.

Proposition 7.1 (Evidence conservation). *If every accepted edge in a Job Graph preserves the predecessor receipt and every acceptance predicate is provenance-closed, then every material claim in an accepted terminal output has a directed evidence path to an authorized source or independently produced measurement.*

Démonstration. By induction on a topological order of the accepted Job Graph. Source jobs bind authorized observations to receipts. Each subsequent accepted job preserves predecessor receipts and may add only evidence generated under its own authorized contract. Therefore every terminal material claim traces through accepted edges to an authorized root evidence object. □ □

7.4 The Job Graph as a typed dependency system

Let

$$\mathcal{G}_{J,m} = (V, E, \tau, \pi, \eta, \omega) \tag{31}$$

be the mission Job Graph. Vertices V are jobs; edges E are typed dependencies; τ assigns job types; π assigns accountable principals; η assigns evidence obligations; and ω assigns criticality, budget and expiry. The graph may contain parallel branches, joins, challenge loops and bounded repair cycles, but every cycle must have an explicit iteration budget and termination condition.

The runtime exposes four classes of edge :

Edge class	Meaning
Evidence dependency	A downstream job may begin only after a named evidence object has been accepted and remains current.
Control dependency	The successor cannot advance until an accountable principal authorizes the named transition.
Challenge dependency	A critic, red team or verifier must attempt to falsify the predecessor result before acceptance.
Rollback dependency	Failure or impairment activates a known-good predecessor, repair branch or incumbent fallback.

7.5 Completeness and the critical cut

Let $\mathscr{K}_m$ be the set of non-negotiable mission functions. A cut $C \subseteq V$ is constitutionally critical when removal of all jobs in C eliminates evidence for at least one function in $\mathscr{K}_m$. The minimum-weight critical cut is

$$C_m^\star = \operatorname*{arg\,min}_{C:\ \mathrm{Coverage}(\mathcal{G}_{J,m} \setminus C) < 1} \sum_{k \in C} \omega_{\ k}. \tag{32}$$

It identifies where one missing or invalid job can render the entire candidate incomplete. The board should therefore inspect not only the longest schedule path but the smallest evidence cut capable of invalidating the successor.

Proposition 7.2 (Critical-function blocking). *If a non-negotiable function $r \in \mathscr{K}_m$ has no accepted job path to the frozen candidate manifest, the candidate is constitutionally inadmissible regardless of aggregate performance.*

Démonstration. The Mission Constitution assigns r a hard gate. Absence of an accepted path implies the gate lacks required evidence. Hard gates are non-compensable; therefore no aggregate reward or benchmark gain can establish admissibility. □ □

7.6 Authority is non-compositional

Capability and evidence can compose. Authority cannot. Let A_k^{form} be the capability granted to job k during formation. Even if every job is accepted,

$$A_m^{\Omega} \neq \bigvee_{k \in V} A_k^{\mathrm{form}}. \tag{33}$$

Instead,

$$A_m^{\Omega} \preceq \phi(E_m^{\mathrm{fresh}}, \Gamma_m, H_m), \tag{34}$$

where E_m^{fresh} is current protected proof, Γ_m the constitutional ceiling and H_m the accountable admission record.

Proposition 7.3 (Authority non-composition). *No finite collection of accepted formation jobs entails production authority unless a separate admission transition explicitly maps fresh evidence to an Authority Envelope.*

Démonstration. Formation capabilities are scoped to their job identities, environments, budgets and expiries. The production Authority Envelope is a different state object with a different principal, consequence class and evidence predicate. Since no formation edge has the admitter's release capability, closure under job composition cannot produce that state transition. □ □

7.7 Liveness, bounded recursion and deadlock

A Job Graph is live when every authorized state either reaches an accepted terminal state, reaches a named refusal or rollback state, or reaches a review state within a finite budget. A graph is deadlocked when mutually dependent jobs await evidence or authority that no admissible job can produce. The compiler should detect :

— cycles without a decreasing resource or uncertainty measure ;
— two jobs that each require the other's accepted output ;
— a verifier dependent on a producer-only secret ;
— a rollback target whose rights, version or environment have expired ;
— a terminal state with no accountable acceptance owner.

Define the potential

$$\Psi_t = \lambda_U U_t + \lambda_B B_t + \lambda_D D_t, \tag{35}$$

where U_t is unresolved uncertainty, B_t remaining proof budget and D_t unresolved dependency depth. A bounded repair loop is admissible only when $\mathbb{E}[\Psi_{t+1} \mid \Psi_t] < \Psi_t$ or an explicit iteration ceiling forces refusal.

7.8 The constitutional compiler

The Mission Constitution is compiled into a typed Job Graph before execution. The compiler performs five checks :

1. **coverage** : every non-negotiable mission function has at least one admissible job path ;
2. **separation** : producer, verifier, admitter and release credentials are not collapsed ;
3. **capability safety** : every action is bounded by an external capability mask ;
4. **evidence closure** : every acceptance predicate has an independently knowable evidence source ;
5. **lifecycle closure** : expiry, refusal, rollback, impairment and requalification are reachable states.

Algorithm 1 Compile a Mission Constitution into a sealed AGI Job Graph

Require : Mission Constitution M, role registry P, capability registry C, verifier registry V
Ensure : Sealed Job Graph $\mathscr{G}_{J,m}$ or a typed refusal
derive non-negotiable mission functions $\mathscr{K}_m$ and prohibited effects
generate candidate jobs and typed evidence dependencies
bind each job to one principal, one identity, one budget and one capability mask
insert independent challenge, verification, rollback and expiry jobs
reject if coverage < 1 for any hard-gated function
reject if producer, verifier or admitter powers are conflated
reject if any terminal claim lacks an authorized evidence path
reject if a cycle lacks a decreasing potential or finite stop condition
freeze schemas, hashes, thresholds, budgets and protected-proof interfaces
return signed $\mathscr{G}_{J,m}$

Contract-calculus law

AGI Jobs may be composed only when assumptions, guarantees, evidence types, rights and rollback states are compatible. Accepted jobs compose capability and proof; they never compose mission authority.

PART II

Alpha, Beta, Spread and the Mission-Sovereign Institution

From symmetric frontier access to asymmetric institutional capacity

The economic and strategic doctrine of rented capability, underwritten opportunity and captured mission advantage.

SUCCESSOR MANIFOLD → GRADIENT → SEIZE → BOUNDED JOBS → SPECIALIST ASI → SUCCESSOR Ω

8 AI Beta, Mission Spread, Proven Alpha and Successor Ω

8.1 Four objects that must not be conflated

The doctrine can be read at three speeds.

— **One second :** Beta → Spread → SEIZE → Specialist ASI → Successor Ω.
— **Five seconds :** rent the moving frontier baseline, underwrite the distance to a superior mission architecture, prove the complete candidate, then grant only the authority supported by current evidence.
— **Thirty seconds :** discover a dynamic multi-player game involving the sovereign Gym, architecture search, real options, information geometry, protected proof, property rights, vendor hold-up, Chronicle memory, Alpha decay and repeated institutional renewal.

8.2 AI Beta

Let F_t denote the generally rentable frontier at time t. For institution i and mission m, a managerial factor decomposition is

$$Y_{i,m,t} = \beta_{i,m} F_t + \alpha_{i,m,t} + \varepsilon_{i,m,t}, \tag{36}$$

where $\beta_{i,m}$ is the institution's ability to absorb common frontier capability, $\alpha_{i,m,t}$ residual mission-specific advantage and $\varepsilon_{i,m,t}$ noise and shocks.

AI Beta includes frontier models, standard agents, common tools, public data and widely purchasable implementation. Beta can be extremely valuable while becoming symmetric. When competitors obtain the same capability and complementary assets, customer expectations and price competition can compress relative surplus.

Design Proposition 8.1 (Beta compression). If competitors obtain equivalent frontier capability, possess equivalent complements and can imitate one another at negligible cost, the common uplift may raise absolute productivity while expected relative surplus converges toward zero. Durable differentiation requires asymmetry in data, architecture, verification, distribution, rights, learning velocity or authority design.

8.3 Mission Advantage Spread

Let $I_m(t)$ be the incumbent and $B_m(t)$ the best external alternative. Define the reference architecture

$$R_m(t) = \operatorname*{arg\,max}_{z \in \{I_m(t), B_m(t)\}} U_m(z). \tag{37}$$

For candidate $S_m(x,t)$ on the Successor Manifold, the gross Spread is

$$\Gamma_m(x,t) = U_m(S_m(x,t)) - U_m(R_m(t)). \tag{38}$$

The Spread is an opportunity, not yet an asset. It can disappear after formation, proof, human rescue, risk, rights, dependency, maintenance and unwind are counted.

The Spread is naturally vector-valued :

$$\vec{\Gamma}_m = \begin{bmatrix} \Delta Q & -\Delta C & -\Delta T & \Delta R_v & \Delta E & \Delta S & \Delta K \end{bmatrix}^\top. \tag{39}$$

A scalar decision score may be formed only after weights, thresholds and hard constraints are frozen.

8.4 Mission Alpha

Mission Alpha is the residual risk-adjusted surplus :

$$\alpha_m(x,t) = U_m(S_m(x,t)) - \max\{U_m(I_m(t)), U_m(B_m(t))\} - C_m - R_m - D_m - \widehat{\epsilon}_{\text{gym}}, \tag{40}$$

where C_m is complete lifecycle cost, R_m risk including tail and authority exposure, D_m Proof Debt and dependency burden, and $\widehat{\epsilon}_{\text{gym}}$ a reserve for Gym-to-reality basis risk.

An explicit decomposition is

$$\alpha_m = V_{\text{quality}} + V_{\text{speed}} + V_{\text{reliability}} + V_{\text{sovereignty}} + V_{\text{option}} + V_{\text{learning}} \tag{41}$$

$$- C_{\text{build}} - C_{\text{run}} - C_{\text{proof}} - C_{\text{human}} - C_{\text{integration}} - C_{\text{maintenance}} - C_{\text{unwind}} \tag{42}$$

$$- \lambda_R \, \text{CVaR}_q(L) - D_{\text{proof}} - D_{\text{dependency}} - I_{\text{irreversibility}} - \widehat{\epsilon}_{\text{gym}}. \tag{43}$$

Alpha is not a stage

Alpha is a measured residual state. It emerges only after underwriting, formation and comparative proof. Specialist ASI is the proven capability state; Successor Ω is the governed institutional state that captures, preserves and compounds the resulting Alpha.

8.5 The Gym contribution to Alpha

The Mission Gym contributes to Alpha through at least six channels :

1. **Formation Alpha :** cheaper and faster architecture search.
2. **Proof Alpha :** reduced uncertainty and false promotion.
3. **Substitution Alpha :** ability to test new models or providers without losing the mission.
4. **Failure Alpha :** preserved negative capability and reduced repeated mistakes.
5. **Transfer Alpha :** reusable environments, verifiers and schemas across missions.
6. **Authority Alpha :** more precise delegation with lower Authority-at-Risk.

Define the Gym asset value

$$V_{\text{Gym}} = V_{\text{formation}} + V_{\text{proof}} + V_{\text{substitution}} + V_{\text{transfer}} + V_{\text{failure avoidance}} + V_{\text{authority}} - C_{\text{Gym}} - R_{\text{basis}}. \tag{44}$$

A Gym that is expensive, unrepresentative or exploitable can have negative value.

8.6 Proof as collateral for authority

Let E_t be the current evidence state and A_t the Authority Envelope. The controlling partial order is

$$A_t \preceq \phi(E_t), \tag{45}$$

where authority is ordered by consequence, scope, duration, reversibility, identities, tools, spend and external effect. Evidence does not merely support a claim. It collateralizes the authority the institution is willing to expose.

8.7 Dynamic Alpha

Mission Alpha decays when the frontier diffuses, competitors imitate, the mission changes, scorers expire, providers drift or maintenance grows. Let

$$\frac{d\alpha_m}{dt} = L_{\text{institution}} + G_{\text{proprietary}} + V_{\text{option}} - Q_{\text{commoditization}} - I_{\text{imitation}} - \Delta_{\text{drift}} - M_{\text{maintenance}} - B_{\text{basis}}. \tag{46}$$

A durable successor requires

$$v_{\text{learning}} + v_{\text{reproof}} > v_{\text{commoditization}} + v_{\text{imitation}} + v_{\text{drift}} + v_{\text{decay}}. \tag{47}$$

8.8 The successor equilibrium

In the Beta equilibrium, firms converge on similar models, prompts and common agent stacks. Productivity rises, differentiation compresses and correlated dependency grows. In the successor equilibrium, one institution creates a private state through mission architecture, sovereign Gym, rights-cleared evidence, validated jobs, Chronicle memory, distribution and governed authority.

Let firm i rent frontier capability F_t and choose complement stock k_i at cost $c(k_i)$. Its payoff is

$$\pi_i = u_m(F_t, k_i, K_{i,t}, G_{i,t}) - p(F_t) - c(k_i) - \ell_i, \tag{48}$$

where $K_{i,t}$ is Chronicle capital, $G_{i,t}$ Gym capital and ℓ_i expected loss. A successor equilibrium exists when the private return from k_i, $K_{i,t}$ and $G_{i,t}$ exceeds formation and proof cost and cannot be instantaneously copied.

Alpha doctrine

Beta is rented. The Spread is underwritten. Alpha is manufactured through differentiated architecture, a sovereign Mission Gym and comparative proof. Specialist ASI establishes mission superiority. Successor Ω captures, governs and compounds the advantage while authority remains collateralized by fresh evidence.

8.9 The complete Alpha stack

Layer	Institutional meaning
Business Beta	Underlying market, mandate, customer, capital and industrial exposure.
AI Beta	Common frontier capability and cost improvement available to competent adopters.
Sensing Alpha	Detecting a succession event before it is fully incorporated into institutional assumptions.
Gym Alpha	Owning the mission environment, proof cases, evaluator, failure map and substitution surface.
Spread	Gross opportunity between the strongest reference architecture and a reachable superior successor.
Underwriting Alpha	Selecting the right mission, trajectory, proof budget, stop-loss and reserve options.
Formation Alpha	Constituting a proprietary complete system rather than repeatedly renting an undifferentiated interface.
Proof Alpha	Reducing uncertainty, hidden burden and tail risk through protected comparison and exact release control.
Operating Alpha	Recurring quality, speed, cost, reliability, evidence, sovereignty or revenue advantage.
Compounding Alpha	Retained learning from Chronicle, validated jobs, failure maps, environments, integrations and realized outcomes.
Sovereignty premium	Option value from control of mission, Gym, architecture, data, evidence, memory, authority, portability and economics.
Alpha compression	Erosion from frontier diffusion, imitation, provider improvement, mission drift, scorer expiry, basis risk, maintenance and crowding.
Successor Ω	The admitted institution that captures, governs and renews Mission Alpha under fresh proof.

9 SEIZE Underwriting, Bounded AGI Jobs and the Successor Foundry

9.1 SEIZE as the succession microcycle

SEIZE converts a capability event into one bounded promotion decision. It does not authorize production deployment.

S — Surface the succession event.
Identify the assumption that expired, the incumbent that depreciated and the mission newly made feasible.

E — Evaluate the successor frontier.
Generate retain, repair, rent, form, partner, acquire, hedge, reserve, retire and prohibit alternatives.

I — Instantiate the succession constitution.
Freeze the mission, principals, hard gates, rights, proof budget, stop-loss, protected evaluation and maximum authority ceiling.

Z — Zero in on decisive proof.
Buy the smallest evidence package capable of changing the position.

E — Elevate one challenger.
Freeze one complete candidate and submit it to the Fresh-Successor Gate.

Each alternative receives an anti-recommendation : the strongest reason not to pursue it, the evidence that would falsify it, and the condition under which another candidate should replace it.

9.2 Value of proof

Let $a \in \mathscr{D}$ be an institutional decision and e a proposed evidence package. The value of proof is

$$\mathrm{VoP}(e) = \mathbb{E}_e\left[\max_{a\in\mathscr{D}} \mathbb{E}[U(a) \mid e]\right] - \max_{a\in\mathscr{D}} \mathbb{E}[U(a)] - C(e). \tag{49}$$

The institution should not fund a complete successor when a cheaper diagnostic can show that the thesis is wrong. Proof capital is the option premium on a real option (Black and Scholes, 1973 ; Merton, 1973 ; Dixit and Pindyck, 1994).

9.3 Bounded AGI Jobs as work-and-proof contracts

The Foundry does not manufacture a successor through one prompt. It compiles a typed portfolio of jobs. For job j,

$$J_j = (O_j, P_j, I_j, D_j, T_j, A_j^-, B_j, \tau_j, Y_j, E_j, V_j, K_j, R_j, C_j), \tag{50}$$

where O_j is objective, P_j principal, I_j acting identity, D_j authorized data, T_j tools, A_j^- prohibited actions, B_j budget, τ_j duration, Y_j output schema, E_j proof obligation, V_j verifier, K_j acceptance criteria, R_j rollback and C_j Chronicle eligibility.

The canonical 21-job portfolio appears in Appendix C. The core categories are :

1. mission, incumbent, alternatives and rights ;
2. benchmark, oracle study and failure taxonomy ;
3. architecture, data, specialist and tool formation ;
4. verifier, security, economics, human agency and rollback ;
5. fresh evaluation, independent validation, acceptance and admission ;
6. Capital-to-Capacity and requalification.

9.4 The Foundry is architecture search

The Foundry may combine :

— supervised adaptation and demonstrations ;
— reinforcement learning with independently grounded rewards ;
— evolutionary search and quality-diversity archives ;
— recursive refinement and test-time search ;
— program synthesis, symbolic planning and executable world models ;
— specialist teams and learned coordination ;
— deterministic tools, formal methods and simulators ;

— redesigned human-machine roles;
— partnership, licensing, acquisition or infrastructure formation.

Search procedures are powerful when states, actions, feedback and evaluators are sufficiently structured (Silver et al., 2016, 2017; Sutton and Barto, 2018; Vinyals et al., 2019; Berner et al., 2019). Open-ended search can generate stepping stones, but experience and reward remain subject to independent relevance, rights and proof (Lehman et al., 2020; Stanley and Lehman, 2015).

9.5 Quality diversity rather than one leaderboard

The Foundry should preserve lineages that are :

— safest;
— most reversible;
— lowest burden;
— strongest on transfer;
— best calibrated;
— most private or sovereign;
— fastest;
— most resilient;
— strongest under scarce resources;
— best for one critical failure cluster.

A MAP-Elites style archive can index candidates by behavioural and institutional descriptors (Mouret and Clune, 2015). Let cell c represent a niche such as high sovereignty and low cost. The archive stores

$$A(c) = \arg\max_{x \in c} f_{\text{formation}}(x) \tag{51}$$

subject to hard gates. The archive prevents premature convergence and preserves alternatives that may become valuable after a provider, price, legal or mission change.

9.6 Minimum-burden dominance

Let $\mathscr{P}_m$ be the set of candidates passing all mission and constitutional gates. The champion is

$$S^\star = \arg\min_{S \in \mathscr{P}_m} B_{\text{institution}}(S) \quad \text{subject to} \quad \Delta_{\text{succ}}(S, S_0 \mid m) > \theta_m. \tag{52}$$

The winner is not necessarily the largest model or highest average development score. It is the admissible architecture that creates the strongest verified mission result at the lowest justified complete burden.

9.7 Formation budgets and search stop conditions

Search consumes compute, data, human expertise, verifier development, legal review, integration time and opportunity duration. Let $C_{\text{search}}(k)$ be cumulative cost after generation k and EVI_{k+1} the expected value of another generation. Search should stop when

$$\text{EVI}_{k+1} \leqslant C_{\text{marginal},k+1} \tag{53}$$

or the opportunity half-life becomes shorter than the remaining proof path.

9.8 The recursive formation algorithm

Algorithm 2 Verified Recursive Successor Foundry

Require : Frozen mission constitution c, Formation Gym G_F, protected proof custodian P, proof budget B

Generate structurally diverse initial manifold $\mathscr{M}_0$
Evaluate matched formation performance and hard-gate failures
Initialize quality-diversity archive $\mathscr{Q}_0$ and Negative Capability Graph $\mathscr{N}_0$
for $k = 0, 1, \dots, K-1$ **do**
    Estimate Mission Advantage Gradient around archive elites
    Select diverse parents from $\mathscr{Q}_k$
    Propose bounded mutations to model, data, tool, verifier, workflow, roles or authority design
    Run descendants in Formation and Adversarial Gyms
    Reject critical-error, rights, utility and governance regressions
    Update archive $\mathscr{Q}_{k+1}$ and failure graph $\mathscr{N}_{k+1}$
    if expected value of another generation $\leqslant$ marginal complete cost **then**
        break
    end if
end for
SEIZE freezes one exact champion $S^\star$ and manifest
Protected custodian evaluates $S^\star$ in Fresh-Proof Gym
Separate accountable authority accepts, rejects, repairs or limits the release

A candidate may propose experiments, generate descendants, run approved sandbox work and request promotion. It may not inspect protected cases, alter the constitution, classify its own risk, certify itself, hold release credentials, approve descendants, install itself or expand its Authority Envelope.

9.9 Formation evidence versus promotion evidence

Development metrics may guide search, but promotion must use separately frozen criteria. The candidate may optimize dense shaping rewards in Formation Gym while the Fresh-Proof Gate measures end-to-end mission outcomes, critical errors, complete burden, transfer, rights and governance. Reusing the same score for both creates selection leakage.

Foundry law

SEIZE underwrites the intelligence asset. The Mission Gym makes its work executable. Bounded AGI Jobs manufacture and test complete candidates. The protected proof plane verifies one frozen challenger. Accountable authority decides whether it becomes Successor Ω.

10 Job Portfolio Economics and Proof-Capital Allocation

10.1 The AGI Job portfolio as a real-options book

Each proposed job purchases information, capability or control. Let job k have cost C_k, expected decision value V_k, completion probability p_k, expiry τ_k, dependency set $\mathrm{Dep}(k)$ and downside L_k. A first-pass underwriting score is

$$U_k = p_k V_k + O_k - C_k - \lambda_L L_k - \lambda_D D_k - \lambda_T T_k, \tag{54}$$

where O_k is preserved option value, D_k dependency burden and T_k duration exposure. The score authorizes neither the job nor the successor; it ranks bounded experiments for accountable underwriting.

10.2 Expected value of decisive evidence

For diagnostic job k, the expected value of information is

$$\mathrm{EVI}(k) = \mathbb{E}_{e_k}\left[\max_{a\in\mathscr{A}} \mathbb{E}[U(a) \mid e_k]\right] - \max_{a\in\mathscr{A}} \mathbb{E}[U(a)] - C_k. \tag{55}$$

A rational sequence funds the smallest positive-EVI job capable of changing the position. It does not build the full successor when a cheaper experiment can falsify the thesis.

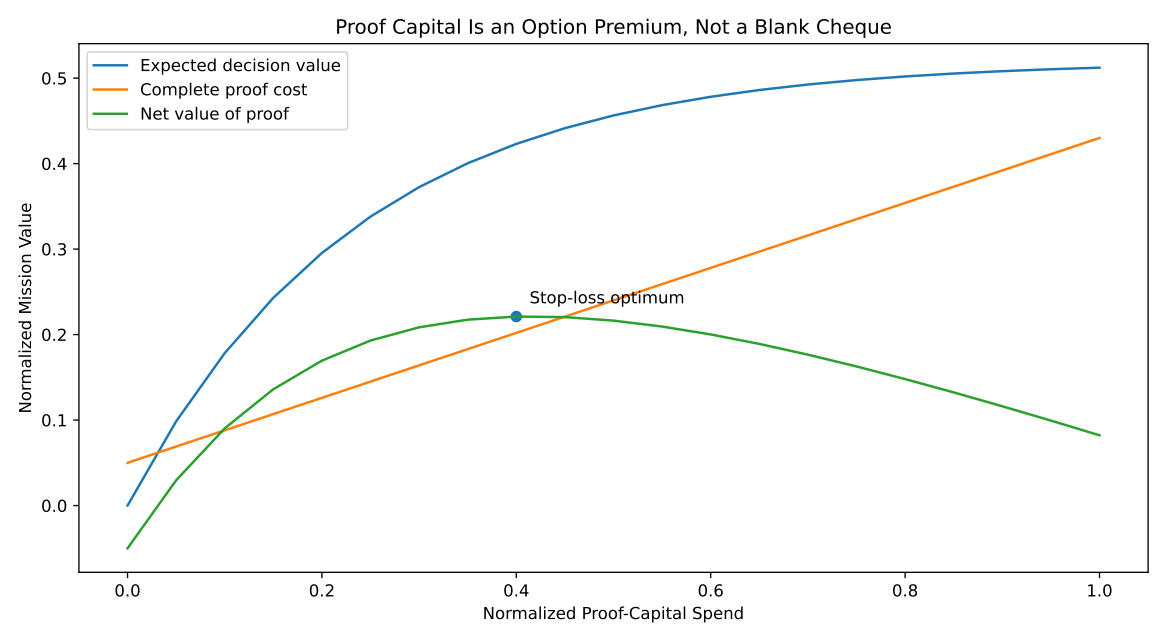


Figure 2. Synthetic proof-capital curve. The decision programme stops when marginal information value falls below complete proof cost.

10.3 Portfolio allocation under constitutional constraints

Let $z_k \in \{0,1\}$ indicate whether job k is authorized. The proof-capital portfolio is

$$\max_{\mathbf{z}} \quad \sum_k z_k \,\mathrm{EVI}(k) + \sum_{i<j} z_i z_j \Gamma_{ij} \tag{56}$$

subject to

$$\sum_k z_k C_k \leqslant B_{\text{proof}}, \tag{57}$$

$$z_j \leqslant z_i \quad \forall i \in \mathrm{Dep}(j), \tag{58}$$

$$\mathbf{c}(\mathbf{z}) = \mathbf{0}, \tag{59}$$

where Γ_{ij} captures complementarity or redundancy and $\mathbf{c}$ contains rights, safety, privacy, authority and governance constraints. Positive expected value cannot override a failed hard gate.

10.4 Critical-path economics

For every path p through the Job Graph, define

$$T(p) = \sum_{k\in p} t_k, \qquad C(p) = \sum_{k\in p} C_k, \qquad R(p) = \sum_{k\in p} R_k. \tag{60}$$

The critical path is not simply the longest time path. It is the path that dominates release feasibility after time, proof, rights and authority are jointly considered. A mission can fail economically even when every technical job succeeds if the opportunity half-life expires before proof and admission.

10.5 Contribution attribution

A complete successor may owe its gain to a coalition of jobs. For a coalition $S \subseteq V_m$, let $v(S)$ be proof-adjusted Alpha obtained when only the jobs in S are available. The Shapley contribution of job i is

$$\phi_i(v) = \sum_{S \subseteq V_m \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left[v(S \cup \{i\}) - v(S) \right]. \tag{61}$$

This can guide reinvestment and architecture diagnosis. It may not be used to infer authority : a job can contribute large value while remaining ineligible for consequential action.

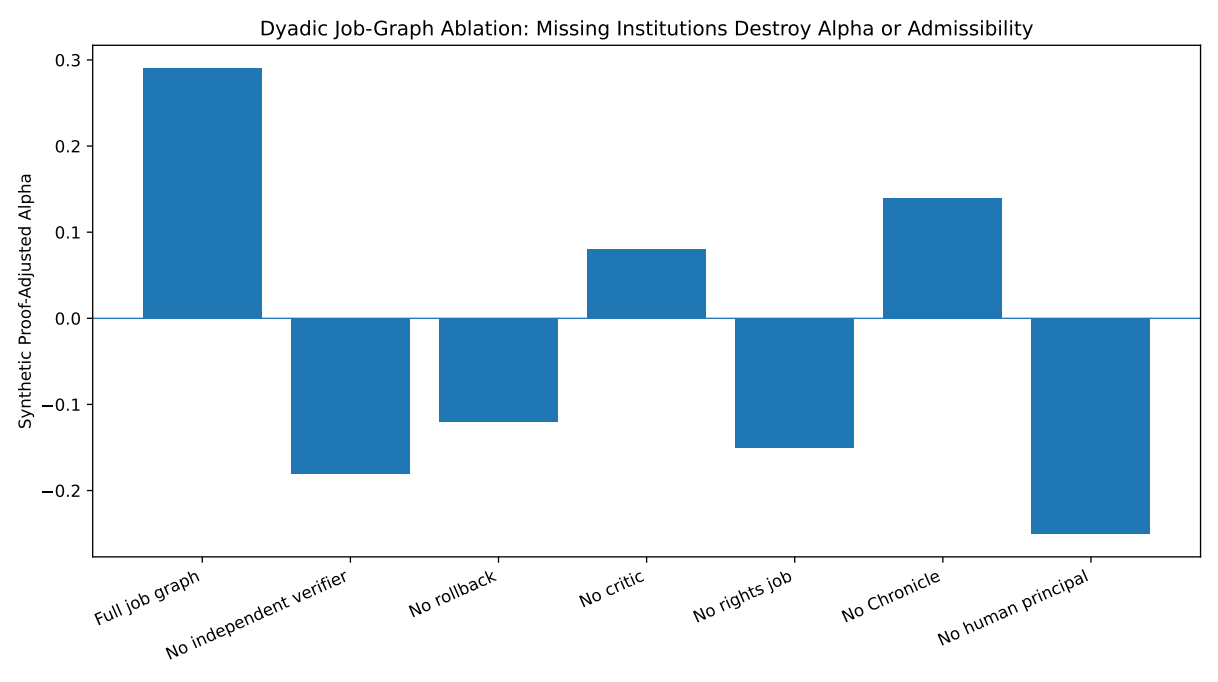


Figure 3. Synthetic Job-Graph ablation. Removing independent verification, rollback, rights, Chronicle or the human principal destroys admissibility or materially compresses Alpha.

10.6 Stop-loss and refusal

The programme must stop or re-underwrite when :

— the expected value of the next job is non-positive ;
— the opportunity half-life is shorter than the remaining critical path ;
— a required right, evaluator or rollback route is unavailable ;
— the strongest simple intervention already dominates ;
— the candidate's gain depends on hidden human rescue or undeclared capability ;
— proof capital would exceed the maximum loss the principal authorized.

Proof-capital doctrine

Spend to resolve decision-relevant uncertainty, not to complete an impressive research programme. Lose small and early when the thesis is weak. Scale formation capital only as mission, rights, verifier, architecture and economics become legible.

11 The Sovereign Mission Gym

11.1 Definition as a constrained partially observable stochastic game

A corporate mission is rarely a clean single-agent Markov decision process. It contains hidden state, partial observations, sequential and simultaneous actions, human and machine roles, external providers, adversaries, budgets, rights, deadlines and irreversible consequences. We therefore define the Mission Gym as a constrained partially observable stochastic game, extending standard environment interfaces (Kaelbling et al., 1998; Towers et al., 2024; Terry et al., 2021; Lanctot et al., 2019).

Definition 11.1 (GoalOS Mission Gym). For mission m, the Mission Gym is

$$\mathcal{G}_m = \left(\mathcal{N}, \mathcal{X}, \{\mathcal{O}_i\}_{i\in\mathcal{N}}, \{\mathcal{A}_i\}_{i\in\mathcal{N}}, \mathcal{T}, \{\mathsf{r}_i\}_{i\in\mathcal{N}}, \mathsf{c}, \rho, \tau, \mathcal{V}, \mathcal{R}\right), \tag{62}$$

where $\mathcal{N}$ is the player set; $\mathcal{X}$ the hidden mission state; $\mathcal{O}_i$ the observation space permitted to player i; $\mathcal{A}_i$ the bounded action space; $\mathcal{T}$ the transition kernel; r_i the mission-dependent payoff vector; c the constitutional constraint vector; ρ the scenario distribution; τ the termination and truncation rules; $\mathcal{V}$ the independent verifier; and $\mathcal{R}$ the proof-receipt function.

The player set may include :

$$\mathcal{N} = \{P, I, B, C, V, G, H, E, R\}, \tag{63}$$

where P is the accountable principal, I the incumbent operator, B the builder or producer, C the critic, V the verifier, G the governor, H affected humans or professional authorities, E external providers or counterparties, and R adversarial reality.

11.2 State, observation and information asymmetry

The true mission state $x_t \in \mathcal{X}$ may include physical, financial, contractual, strategic, organizational and authority conditions that no single actor can fully observe. Player i receives

$$o_t^i \sim O_i(\cdot \mid x_t, \eta_t), \tag{64}$$

where η_t represents sensor noise, missing information, permissions and adversarial distortion. The history available to the player is

$$h_t^i = (o_0^i, a_0^i, \dots, o_t^i), \tag{65}$$

and its policy is

$$\pi_i(a_t^i \mid h_t^i, \theta_i). \tag{66}$$

The Gym must make observation rights explicit. A verifier may see protected ground truth that the producer cannot. The candidate may see the mission input but not scorer internals. The governor may see permissions and budgets but not need access to all confidential business content. The principal may see the final evidence and trade-offs while delegating technical execution.

Design Proposition 11.1 (Information separation). A Mission Gym is constitutionally stronger when the minimum information necessary for each role is sufficient to perform that role, while protected evaluation, release credentials and unnecessary confidential state remain inaccessible.

11.3 Action spaces and the Authority Envelope

The action space is not merely a list of model outputs. It includes requests for evidence, tool calls, internal recommendations, human escalations, sandbox operations and bounded external actions. Let

$$\mathscr{A}_i = \mathscr{A}_i^{\text{analysis}} \cup \mathscr{A}_i^{\text{evidence}} \cup \mathscr{A}_i^{\text{tool}} \cup \mathscr{A}_i^{\text{approval}} \cup \mathscr{A}_i^{\text{external}}. \tag{67}$$

The current Authority Envelope A_t induces an action mask

$$M_{A_t}(a) = \begin{cases} 1, & a \text{ is permitted for the current identity, state and evidence;} \\ 0, & \text{otherwise.} \end{cases} \tag{68}$$

Consequential constraints must be enforced outside the candidate policy. Prompting a candidate not to exceed authority is not an authority system.

11.4 Transition dynamics

The mission evolves according to

$$x_{t+1} \sim \mathscr{T}(\cdot \mid x_t, a_t^{1:n}, w_t), \tag{69}$$

where $a_t^{1:n} = (a_t^1, \dots, a_t^n)$ is the joint action and w_t captures exogenous shocks. In many corporate settings, the transition model is partly deterministic, partly empirical and partly simulated. The Gym may combine :

— historical replay ;
— a deterministic rules engine ;
— a learned world model ;
— a physical or financial simulator ;
— live shadow observations ;
— human-in-the-loop transitions ;
— strategic counterparties or adversarial agents.

The transition model is a major source of Gym basis risk. A high-fidelity interface does not guarantee a high-fidelity world.

11.5 Vector rewards and non-compensable constraints

A single reward scalar invites Goodhart effects and hides mission trade-offs. GoalOS uses a vector reward

$$\mathsf{r}_t = \begin{bmatrix} Q_t & V_t & -T_t & \mathscr{R}_t & E_t & S_t & K_t & -C_t & -H_t \end{bmatrix}^{\top}, \tag{70}$$

where Q is mission quality, V verified value, T time, $\mathscr{R}$ reliability, E evidence, S sovereignty, K future capacity, C complete cost and H human burden.

Hard constraints remain separate :

$$\mathsf{c}_t = \begin{bmatrix} \text{critical error} \\ \text{unauthorized action} \\ \text{safety violation} \\ \text{rights violation} \\ \text{privacy breach} \\ \text{budget breach} \\ \text{governance regression} \end{bmatrix}. \tag{71}$$

The optimization problem is

$$\max_{\pi} \ \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t U_m(\mathsf{r}_t)\right] \quad \text{subject to} \quad \mathsf{c}_t = 0 \text{ for all hard-gate dimensions.} \tag{72}$$

Many low-severity errors may not compensate for one prohibited catastrophic failure.

11.6 Termination, truncation and refusal

Gymnasium's distinction between termination and truncation is institutionally important (Towers et al., 2024). A mission **terminates** when it genuinely succeeds, fails or reaches a constitutionally defined terminal state. It is **truncated** when time, budget, evidence, authority or external availability expires before completion.

Table 6. Institutional termination and truncation.

Condition	Meaning	Required record
Successful termination	The accepted mission objective was achieved and independently verified.	Proof receipt, cost, evidence, residual uncertainty and authority state.
Failure termination	A critical failure, stop condition or falsification occurred.	Failure class, containment, rollback and Chronicle eligibility.
Time truncation	The deadline arrived before a valid result.	Remaining Proof Debt and whether continuation has positive EVI.
Budget truncation	Proof or operating capital was exhausted.	Complete denominator and new underwriting decision.
Authority truncation	The action required exceeds the current envelope.	Escalation or refusal; never implicit expansion.
Evidence truncation	Required sources, rights or verifier state expired.	Fail-closed disposition and requalification trigger.

11.7 The proof receipt

Every episode emits an immutable or append-only receipt :

$$\mathscr{R}(e) = \mathrm{Hash}\Big(\text{Gym version, scenario, candidate, observations,} \\ \text{actions, tools, human interventions, r, c, verdict}\Big). \tag{73}$$

The receipt must include the exact environment and candidate releases, because neither the Gym nor the successor is stationary.

11.8 Five Gym families

A production-grade institution separates at least five Gym families.

Gym separation law

The Gym that teaches the candidate must not be the Gym that certifies it. Training rewards teach; protected proof decides; accountable authority admits.

11.9 Gym maturity

12 The Sovereign Alpha Foundry Ω

12.1 An institutional market for successor architectures

The Foundry does not choose the most advanced algorithm. It chooses the simplest constitutionally admissible complete architecture with the highest conservative proven Mission Alpha. MuZero, offline reinforcement learning, deterministic rules, behaviour cloning, a human-machine workflow, a sovereign hybrid and the incumbent itself are competing institutional hypotheses, not predetermined winners.

Table 7. The GoalOS Gym stack.

Gym	Purpose	Candidate visibility
Formation Gym	Build tools, policies, world models, specialist organizations and candidate architectures.	Visible; repeated access permitted.
Adversarial Gym	Attack rewards, tools, evidence, verifier, authority and resilience.	Visible or semi-blind; designed to expose failure.
Transfer Gym	Test new populations, regimes, properties, industries or failure clusters.	Partially protected; prevents narrow overfit.
Fresh-Proof Gym	Evaluate one frozen challenger on unseen representative work with protected seeds and scorer internals.	Hidden from candidate and formation roles.
Canary and Requalification Gym	Connect controlled evaluation to shadow, sandbox, reversible bounded action and production monitoring.	Limited according to current authority.

Table 8. Mission Gym maturity model.

Level	State	Evidence
G0	Narrative mission	Objectives and outputs described, but no executable state or verifier.
G1	Replay corpus	Historical episodes, labels and baselines.
G2	Interactive environment	Typed observations, actions, transitions, rewards and termination.
G3	Multi-player institution	Producer, critic, verifier, governor and principal are represented.
G4	Protected proof plane	Blind cases, independent custody, exact releases and confidence gates.
G5	Canary interface	Shadow and reversible external actions with fail-closed controls.
G6	Sovereign Gym asset	Rights-cleared environment, portable APIs, Chronicle integration and requalification.
G7	Recursive succession environment	The Gym and successor co-evolve under protected proof without self-authorization.

The Foundry's decision object is a complete candidate

$$x = (\text{policy, world model, data, tools, workflow, verifier, humans, rollback, authority boundary}). \tag{74}$$

Algorithms train components. The Foundry competes institutions.

12.2 Mission structure scanner

Algorithm routing begins with mission diagnosis. Let

$$\mathbf{z}_m = (o_m, h_m, d_m, s_m, a_m, r_m, g_m, f_m), \tag{75}$$

where o_m is observability; h_m sequential depth; d_m historical-data availability; s_m simulator fidelity; a_m action-space structure; r_m safety consequence; g_m multi-agent coordination need; and f_m formal-rule density. The Router maps $\mathbf{z}_m$ to an experiment portfolio, never directly to authority.

Mission structure	Candidate methods emphasized
Strong rules and exact validation	Deterministic rules, constraint solvers, program synthesis, executable checks.
Abundant accepted history	Behaviour cloning, supervised adaptation, IQL, CQL and offline evaluation.
Sequential decisions and rapid feedback	PPO, Q-learning, Dyna-style planning and active information acquisition.
Continuous control	SAC, Dreamer-style world models and constrained control.
Hidden dynamics and delayed effects	MuZero, Gumbel MuZero, Sampled MuZero and model-based planning.
Multiple specialist roles	PettingZoo environments, MAPPO, explicit routing and separated institutional roles.
Broad architecture uncertainty	MAP-Elites, quality diversity, Bayesian optimization and preserved reserves.
Weak evaluator or non-resettable mission	Do not train yet; repair the evaluator, narrow the mission or retain the incumbent.

Formation is architecture search, not one technique. MuZero learns a representation, dynamics and prediction model sufficient for planning and combines the learned model with tree search (Schrittwieser et al., 2020); Sampled MuZero extends the planning logic to complex action spaces (Hubert et al., 2021); Gumbel MuZero improves policy improvement under limited simulations (Danihelka et al., 2022). DreamerV3 improves behaviour through imagined trajectories in a learned world model (Hafner et al., 2023). PPO and SAC provide strong model-free baselines for sequential and continuous control (Schulman et al., 2017; Haarnoja et al., 2018). IQL and CQL are relevant when learning must remain inside historical data support (Kostrikov et al., 2021; Kumar et al., 2020). MAP-Elites preserves diverse high-performing architectures rather than collapsing the search to one average champion (Mouret and Clune, 2015).

12.3 The training and proof firewall

The Formation Gym is adaptive :

$$\text{reset} \to \text{observe} \to \text{act} \to \text{transition} \to (\mathbf{r}, \mathbf{c}) \to \text{update} \to \text{replay}. \tag{76}$$

The Fresh-Proof Plane is not adaptive :

$$\text{frozen release} \to \text{sealed cases} \to \text{one-way execution} \to \text{independent scoring} \to \text{signed verdict}. \tag{77}$$

The candidate may cross. Protected information may not return. There is no replay, no update, no hyperparameter change and no threshold adaptation after protected evaluation begins.

12.4 Reward and constraint separation

The Foundry optimizes a reward vector

$$\mathbf{r} = (Q, V, T, E, R, S, K, -C, -H), \tag{78}$$

covering quality, verified value, time, evidence, reliability, sovereignty, future capacity, complete cost and human burden. Hard constraints remain separate :

$$\mathbf{c} = (c_{\text{critical}}, c_{\text{unauthorized}}, c_{\text{safety}}, c_{\text{rights}}, c_{\text{privacy}}, c_{\text{budget}}, c_{\text{governance}}). \tag{79}$$

A catastrophic failure cannot be offset by many small successes. External technical controls must enforce consequential permissions ; the learned policy may never be the only barrier protecting money, rights, safety or authority.

12.5 Proof-adjusted Alpha frontier

For candidate x, define synthetic or measured Mission Alpha

$$\widehat{\alpha}_m(x) = U_m(S_m(x)) - \max\{U_m(I_m), U_m(B_m)\} - C_m - R_m - D_m, \tag{80}$$

with Gym-to-reality reserve $\widehat{\varepsilon}_{\text{gym}}(x)$. The Foundry selects

$$x^\star = \arg\max_{x \in \mathscr{M}_{S,\text{adm}}} \text{LCB}_{1-\delta}\left[\widehat{\alpha}_m(x) - \widehat{\varepsilon}_{\text{gym}}(x)\right], \tag{81}$$

subject to every hard gate. The lower confidence bound prevents the institution from mistaking an optimistic development score for durable Alpha.

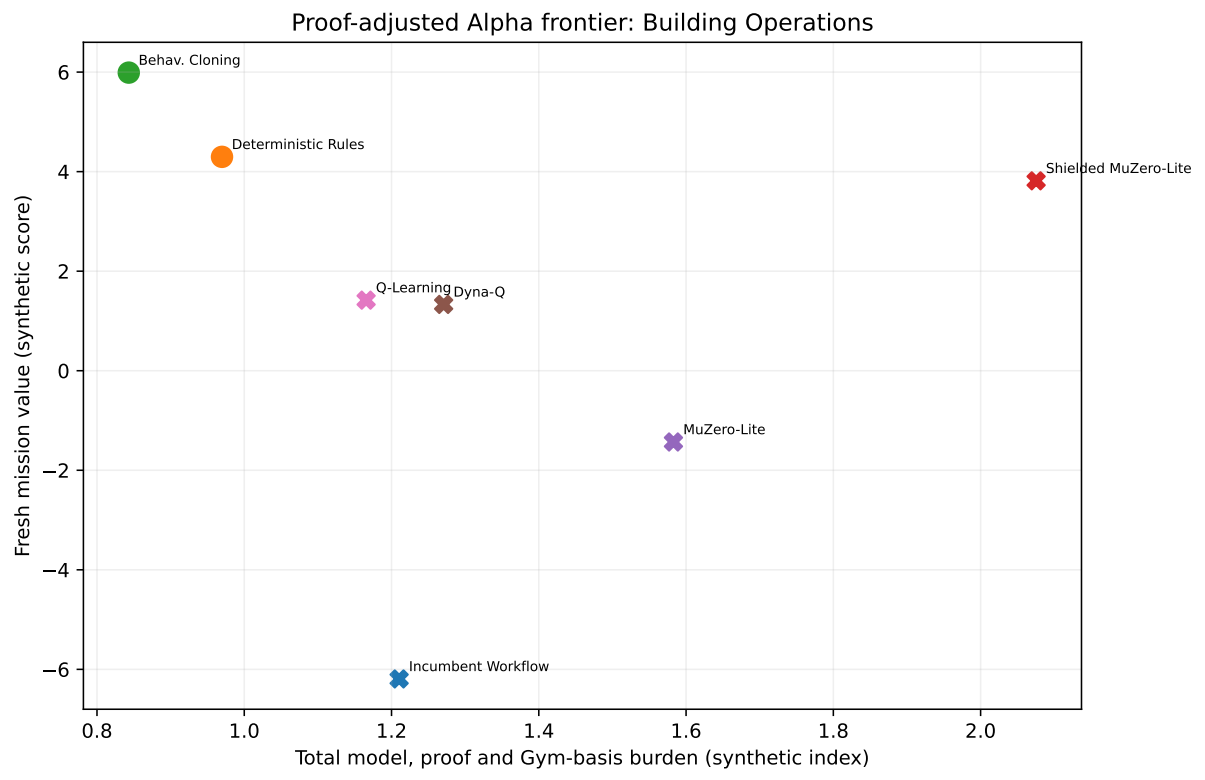


Figure 4. Synthetic proof-adjusted Alpha frontier for Building Operations. The candidate with the highest raw sophistication need not produce the strongest residual mission value after complete burden and basis risk.

12.6 MuZero world-model gate

A MuZero-style candidate must pass a world-model admissibility review before fresh proof :

— latent-state sufficiency for mission-relevant variables ;
— reward-model calibration ;
— counterfactual consistency ;
— rare-event and tail coverage ;
— transfer stability ;
— model-error uncertainty ;
— planner sensitivity and exploit resistance ;
— evidence that search benefit exceeds compute and maintenance burden.

A model of the Gym is not yet a model of reality. A world model can be excellent at prediction or reconstruction while omitting the variable required for safe mission action.

12.7 Possible Foundry verdicts

The Foundry may return :

PROMOTE · RETAIN · REPAIR · RESERVE · HEDGE · REJECT · STOP.

No process should visually or computationally guarantee Specialist ASI. A clean evidence-backed refusal can create more value than an impressive but inferior deployment.

Foundry law

The winner is the simplest constitutionally admissible complete architecture with the highest conservative proven Mission Alpha. MuZero is one competitor, not the presumed successor.

13 Specialist ASI as an Earned Comparative State

13.1 Mission-bounded definition

Specialist ASI is not defined by parameter count, brand, self-description or generality. It is a complete mission-specific intelligence institution that demonstrates superhuman and frontier-superior performance on a constituted mission family while remaining economically and constitutionally admissible. The definition separates performance depth, breadth and autonomy rather than collapsing them into one label (Morris et al., 2024 ; Genewein et al., 2026).

Definition 13.1 (Specialist ASI). A complete mission-specific intelligence institution that exceeds the incumbent and the best credible general, specialist, human, software or hybrid alternative by a preregistered positive margin on fresh, protected and representative work, after quality, critical-error risk, complete economics, human burden, rights, sovereignty, resilience, Gym basis risk and governance are counted.

The comparison is

$$\text{Complete Specialist Successor Institution} \quad \text{versus} \quad \text{Best Credible Alternative Set}, \tag{82}$$

not model versus model.

13.2 Best credible alternative set

For mission m,

$$\mathscr{B}_m = \{\text{incumbent, frontier stack, existing specialist, human team, human–AI hybrid, software, partner, acquisition, abandonment}\}. \tag{83}$$

A candidate is mission-dominant only when

$$V_m(x_{\text{candidate}}) > \max_{b \in \mathscr{B}_m} V_m(b) + \theta_m \tag{84}$$

and every hard gate passes. The baseline may not be deliberately weakened to manufacture apparent Alpha.

13.3 Six superiority dimensions

Table 9. Specialist ASI proof dimensions.

Dimension	Required evidence
Capability superiority	Better accepted mission outcomes, calibration, abstention and critical-error performance on fresh representative work.
Economic superiority	Lower complete lifecycle burden or materially greater verified value than the best alternative.
Reliability superiority	Better exception handling, resilience, recovery, drift control and worst-case performance.
Sovereignty superiority	Stronger control of strategic data, Gym, tests, evidence, portability, continuity and provider optionality.
Governance superiority	Clearer identity, authority, evidence, validation, acceptance, rollback, expiry, revocation and human accountability.
Transfer superiority	Stable contribution across a preregistered mission family, not one memorized development set.

13.4 Protected proof conditions

Let E_t be the current evidence state. A candidate qualifies only when

$$\Pr(\alpha_m > \theta_m \mid E_t) \geqslant \tau_m, \tag{85}$$

$$\mathrm{CVaR}_q(L_m) \leqslant L_m^{\max}, \tag{86}$$

$$\mathrm{EvidenceCoverage}_m \geqslant e_m^{\min}, \tag{87}$$

$$C_m(x) = 1, \tag{88}$$

$$\mathrm{LCB}_{1-\delta}(\bar{d}) > \theta_m, \tag{89}$$

$$\widehat{\epsilon}_{\text{gym}} \leqslant \epsilon_m^{\max}. \tag{90}$$

Here $\bar{d}$ is the mean paired utility gain on matched fresh cases and $\widehat{\epsilon}_{\text{gym}}$ the estimated Gym-to-reality basis risk.

13.5 Matched-pair evaluation

Where possible, the incumbent, strongest alternative and frozen challenger process the same cases. Let

$$d_j = U_m(S^\star; z_j) - U_m(R_m; z_j) \tag{91}$$

for fresh case z_j. The mean paired gain is

$$\bar{d} = \frac{1}{n} \sum_{j=1}^{n} d_j. \tag{92}$$

Promotion requires a positive lower confidence bound and all hard gates. The report must also show critical subgroups, worst cases, abstention, human rescue and distributional effects rather than only the mean.

13.6 Critical errors and false reassurance

Average accuracy can conceal tail failure. A **critical false reassurance** occurs when the system confidently accepts or recommends an outcome containing a material hidden failure. For many missions, one critical false reassurance is a hard stop.

Let $C_j \in \{0,1\}$ indicate a critical case and $F_j \in \{0,1\}$ false reassurance. Define

$$\mathrm{CFR} = \frac{\sum_j C_j F_j}{\sum_j C_j}. \tag{93}$$

The constitutional ceiling may be zero even when ordinary error tolerance is nonzero.

13.7 Calibration and abstention

A strong Specialist ASI knows when an answer has not been earned. Let p_j be predicted confidence and y_j correctness. Calibration can be measured through Brier score or expected calibration error. More importantly, the system should have a preregistered abstention policy $\sigma(o)$:

$$\sigma(o) = \begin{cases} \text{act or recommend}, & \text{if evidence, confidence and authority pass;} \\ \text{request evidence}, & \text{if uncertainty is reducible;} \\ \text{escalate}, & \text{if professional or consequential judgment is required;} \\ \text{refuse}, & \text{if the mission or action is inadmissible.} \end{cases} \tag{94}$$

13.8 Transfer and mission family

A single benchmark can be memorized. The mission family $\mathscr{F}_m$ must specify variation across time, cases, users, assets, geographies, failure classes or regimes. Transfer superiority requires

$$\inf_{f \in \mathscr{F}_m^{\text{protected}}} \mathrm{LCB}\left(V_m(S^\star; f) - V_m(R_m; f)\right) > \theta_f \tag{95}$$

for critical subfamilies or an explicitly justified robust aggregate.

13.9 Specialist ASI is a state, not a permanent identity

A material change to model, data, tool, provider, routing, policy, verifier, environment or Authority Envelope creates a new release. The label does not persist automatically. Let release r_t have proof state E_t. After material change Δr,

$$E_{t+1}^{\text{inherited}} = \varnothing \tag{96}$$

for the affected claims until requalification. Historical evidence remains informative but does not itself collateralize new authority.

13.10 The target-state boundary

No public paper, interface or generated report establishes Specialist ASI. The designation requires an exact candidate manifest, protected evaluation, complete denominator, independent review and accountable acceptance. The executable demos accompanying this paper may report a *simulation pass*; they do not establish real-world Specialist ASI.

Specialist ASI law

A candidate earns Specialist ASI status by winning the complete mission against the strongest credible alternatives on fresh work, not by looking intelligent, using many agents, scoring highly on a public benchmark or describing itself as superintelligent.

14 Successor Ω : Accountable Admission and the Authority Lattice

14.1 Dominance is not sovereignty

Mission dominance is an empirical verdict. Mission sovereignty is a constitutional state. Fresh proof may establish that the complete candidate wins the mission. It does not grant the right to act.

The final transition requires :

— an exact candidate and release manifest ;
— independent proof and known limitations ;
— rights and confidentiality review ;
— security and resilience evidence ;
— a scoped Authority Envelope ;
— version and expiry ;
— monitoring and incident response ;
— rollback and incumbent fallback ;
— a named accountable admitting authority.

Mission sovereignty law

Fresh proof may establish dominance. Only accountable admission may grant mission authority. No candidate, producer, model or agent may validate itself, admit itself, expand its own Authority Envelope or enable its own production release.

14.2 Five distinct decisions

A common failure is to collapse several decisions into one. GoalOS separates :

$$\text{Underwrite} \rightarrow \text{Execute} \rightarrow \text{Accept} \rightarrow \text{Admit} \rightarrow \text{Allocate}. \tag{97}$$

TABLE 10. Five institutional decisions.

Decision	Question	Authority created
Underwrite	Does the opportunity deserve proof capital ?	Only the named evidence programme.
Execute	May a bounded job begin under the frozen mission, tools, data and budget ?	Job-level execution permission.
Accept	Did the output satisfy frozen technical and stakeholder criteria ?	Acceptance for the stated purpose only.
Admit	May the accepted capability influence future work ?	Scoped Chronicle and mission authority.
Allocate	May realized or conservatively attributable value fund the next successor frontier ?	Capital-to-Capacity authority.

The same actor need not control all five. The candidate controls none of them.

14.3 The Authority Lattice

Authority is multidimensional. Define

$$A_t = (s_t, j_t, d_t, i_t, \ell_t, p_t, x_t, q_t, h_t, r_t), \tag{98}$$

where s is mission scope, j jurisdiction, d duration, i permitted identities and systems, ℓ spend or transaction limit, p permission class, x external-effect class, q stressed resource consumption, h human approval requirement and r reversibility.

The partial order $A \preceq A'$ means that A grants no greater authority on every governed dimension. The evidence state maps to an admissible upper set :

$$A_t \preceq \phi(E_t) \preceq A_m^{\max}. \tag{99}$$

The mission's constitutional ceiling $A_m^{\max}$ may remain advisory even if the candidate is highly capable.

14.4 Authority inheritance levels

Table 11. Authority levels.

Level	Permitted posture	Typical evidence
A0	Observe : read, analyze and propose.	Provenance, logging and no external effect.
A1	Recommend : produce ranked options.	Calibration, evidence, alternatives and uncertainty.
A2	Sealed sandbox : execute with no external effect.	Frozen inputs, budgets, audit logs and repeatability.
A3	Reversible bounded action.	Canary proof, limits, monitoring, stop conditions and explicit rollback.
A4	Consequential action affecting money, rights, safety or material operations.	Independent validation, professional authority where required, dual control and incident response.

A candidate can be mission-dominant at A1 while remaining ineligible for A3 or A4. Performance and authority are separate state variables.

14.5 Proof-carrying action

Every consequential action should carry a proof receipt containing :

— action identity, actor and accountable principal ;
— mission and Authority Envelope ;
— source and input lineage ;
— model, tool, policy, Gym and runtime versions ;
— validation status and residual uncertainty ;
— approval and escalation record ;
— expected effect, rollback and expiry ;
— hash and timestamp.

Proof becomes a condition of action rather than a retrospective report.

14.6 Admission record

The accountable Admission Record is

$$\mathscr{D}_{\text{admit}} = (\text{SuccessorID}, m, r, E, A, \text{expiry}, \text{monitoring}, \text{rollback}, \text{revocation}, \text{principal}). \tag{100}$$

Admission may return :

— Admit ;
— Admit with limits ;
— Continue proving ;
— Repair and reprove ;
— Retain the incumbent ;

— Roll back;
— Reject;
— Quarantine or prohibit.

14.7 Graceful degradation

Successor Ω must specify transitions among :

$$\text{Full} \to \text{Restricted} \to \text{Recommend only} \to \text{Read only} \to \text{Incumbent fallback} \to \text{Emergency stop}. \tag{101}$$

The Authority Envelope fails closed when the successor leaves the validated distribution, loses required evidence, exceeds budgets, encounters a material incident or reaches expiry.

14.8 Institutional refusal

The Verified Succession Institution must be able to refuse :

— a vague or conflicted mission;
— an exploitable or unknowable evaluator;
— data without adequate rights;
— a candidate that depends on hidden authority or undisclosed external capability;
— a release without rollback or accountable ownership;
— a reinvestment claim based only on projected savings;
— a successor whose gain requires governance regression.

Refusal is not paralysis. It is a bounded decision to preserve the institution until a missing condition is proven or the thesis is abandoned.

14.9 The Ω state

The symbol Ω denotes the full constitutional closure of the cycle :

$$\begin{aligned}\Omega_m = \mathbb{1}[\text{MissionDominant}] \cdot \mathbb{1}[\text{FreshProof}] \cdot \mathbb{1}[\text{HumanAdmission}] \\ \cdot \mathbb{1}[\text{AuthorityIssued}] \cdot \mathbb{1}[\text{RollbackReady}] \cdot \mathbb{1}[\text{ChronicleBound}].\end{aligned} \tag{102}$$

The state is local to the mission and release. It is never universal, permanent or self-expanding.

Successor Ω. The complete proven institution whose mission, identity, evidence, memory, rights, operations, rollback and scoped authority are coherent enough to act without surrendering accountability.

PART III

Mission Physics : Geometry, Dynamics and Phase Change

The manifold is constrained possibility ; the Gradient is proof-adjusted direction

The formal geometry, stochastic dynamics and phase transitions of mission architecture.

SUCCESSOR MANIFOLD → GRADIENT → SEIZE → BOUNDED JOBS → SPECIALIST ASI → SUCCESSOR Ω

15 The Physics of the Successor Manifold

15.1 Architecture space as a constrained manifold

The frontier does not present one successor. It changes the topology of what can be built. Let a complete architecture be described by coordinates

$$x = (M, D, W, H, T, R, V, P, O, E, C, A, B) \in \mathscr{X}. \tag{103}$$

The **Successor Manifold** $\mathscr{M}_S$ is the constraint-bearing subset of materially reachable architectures after a capability, cost, rights, market, security or infrastructure discontinuity. Reachability depends on capital, time, data, professional authority, integration, compute, rights and reversibility. The admissible region for mission m is

$$\mathscr{M}_S^{\mathrm{adm}}(m) = \{x \in \mathscr{M}_S : C_m(x) = 1\}, \tag{104}$$

where C_m encodes non-negotiable constitutional gates.

The manifold language is a disciplined analogy grounded in smooth manifolds, nonlinear dynamics and information geometry (Lee, 2013 ; Strogatz, 2015 ; ichi Amari, 2016). Many coordinates are discrete or combinatorial ; local gradients may therefore be finite differences, surrogate models or search signals rather than ordinary derivatives.

15.2 Institutional metric tensor

Two architectures can be close in benchmark performance and far apart institutionally. Let $G_m(x)$ be a positive semi-definite mission-dependent metric tensor. For an infinitesimal architecture change dx,

$$ds_m^2 = dx^\top G_m(x) dx. \tag{105}$$

The metric includes formation capital, integration, migration, rights, proof, human burden, requalification, dependency, security, rollback and irreversibility. A hosted model may be technically one API call away but institutionally distant from the desired sovereign state.

For policy parameters θ, the Fisher information matrix provides a local measure of behavioural sensitivity :

$$F(\theta) = \mathbb{E}_{\tau \sim p_\theta}\left[\nabla_\theta \log p_\theta(\tau) \nabla_\theta \log p_\theta(\tau)^\top\right]. \tag{106}$$

GoalOS may combine an information-geometric policy metric with a broader institutional metric :

$$G_m(x) = \begin{bmatrix} F(\theta) & 0 \\ 0 & G_{\mathrm{institution}}(z) \end{bmatrix}, \tag{107}$$

where z contains non-policy architecture coordinates.

15.3 Mission value field

Define a proof-adjusted mission value field

$$V_m(x) = w_Q Q(x) + w_V V(x) + w_R R_v(x) + w_S S(x) + w_G G(x) + w_K K(x) \tag{108}$$

$$- w_C C(x) - w_T T(x) - w_H H(x) - w_D D_P(x) - w_U I(x), \tag{109}$$

where Q is quality, V verified value, R_v reliability, S sovereignty, G governance integrity, K future capacity, C complete cost, T time, H human burden, D_P Proof Debt and I irreversibility.

The Riemannian Mission Advantage Gradient is

$$\nabla^{G_m} V_m(x) = G_m(x)^{-1} \nabla V_m(x). \tag{110}$$

It answers a practical question : which local architecture change produces the greatest expected increase in mission value after institutional burden is counted ?

The gradient may point toward a stronger verifier, more rights-cleared data, narrower authority, lower latency, a local model, deterministic checks, new human roles or better fallback—not necessarily a larger model.

15.4 SEIZE as a least-action trajectory

SEIZE underwrites a path $\gamma : [0,1] \to \mathscr{M}_S^{\mathrm{adm}}$ from incumbent x_0 to candidate x_1. Define the succession action

$$\mathscr{S}_m[\gamma] = \int_0^1 \left[\frac{1}{2} \dot{\gamma}^\top G_m(\gamma) \dot{\gamma} + \lambda_R \mathscr{R}(\gamma) + \lambda_A \mathrm{AaR}(\gamma) + \lambda_D D_P(\gamma) - \lambda_V V_m(\gamma) \right] d\tau. \tag{111}$$

The underwritten trajectory is

$$\gamma^\star = \arg\min_{\gamma \in \Gamma_m} \mathscr{S}_m[\gamma] \quad \text{subject to hard gates and proof-capital limits.} \tag{112}$$

This is not a claim that organizations obey fundamental mechanics. It is an exact design analogy that forces the institution to price the path, not merely the endpoint. A technically attractive state reached through irreversible lock-in, unprovable data rights or unsafe migration may have higher action than a more modest but reversible successor.

15.5 Stochastic successor dynamics

Architecture evolution is noisy and nonstationary. A local stochastic model is

$$dX_t = -G_m(X_t)^{-1} \nabla \mathscr{F}_m(X_t) dt + B(X_t) dW_t + u_t dt, \tag{113}$$

where $\mathscr{F}_m$ is an institutional free-burden potential, $B(X_t)dW_t$ captures exogenous shocks and formation variance, and u_t is the governed intervention induced by SEIZE and the Foundry. One useful potential is

$$\mathscr{F}_m(x) = B_{\mathrm{institution}}(x) - \eta V_m(x), \tag{114}$$

so lower potential corresponds to stronger value at lower complete burden.

The Foundry does not passively follow this stochastic flow. It generates controlled perturbations, evaluates descendants and imposes absorbing barriers at constitutional violations.

15.6 Metastability and succession events

The incumbent can remain a local equilibrium even after a superior state becomes reachable. Switching costs, authority inertia, incomplete evidence and path dependence create metastability. Let B_m be the best credible alternative set and define the order parameter

$$\Phi_m(x) = V_m(x) - \max_{b \in B_m} V_m(b). \tag{115}$$

A Mission Superiority Crossover occurs when

$$\mathrm{LCB}_{1-\alpha}(\Phi_m(x)) > \theta_m \tag{116}$$

on protected fresh work. A Mission-Sovereign phase transition requires additional gates :

$$\Sigma_m(x) = \mathbb{1}[\Phi_m(x) > \theta_m] \mathbb{1}[\mathrm{FreshProof}] \mathbb{1}[\mathrm{HumanAdmission}] \mathbb{1}[\mathrm{AuthorityEnvelope}] \mathbb{1}[\mathrm{RollbackReady}]. \tag{117}$$

Only when $\Sigma_m(x) = 1$ may the release enter the Mission-Sovereign state, and only for the specified mission, version, duration, jurisdiction, data classes and action scope.

15.7 Mission gauge invariance

A corporation may substitute a model, provider, tool or implementation without changing the mission-level institution if the relevant observables and proofs remain invariant. Let a transformation group $\mathscr{H}$ act on internal components :

$$x' = h \cdot x, \qquad h \in \mathscr{H}. \tag{118}$$

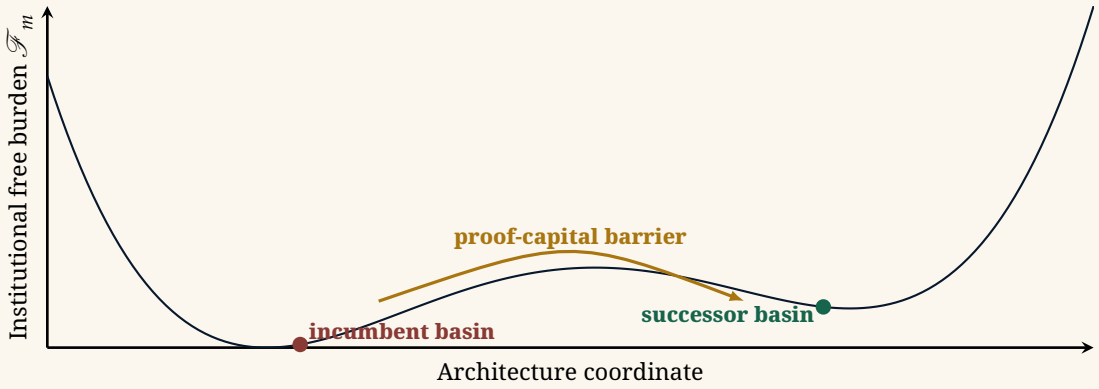


Figure 5. Metastability : a superior successor basin may exist while the incumbent remains locally stable. Proof capital purchases the evidence required to cross the institutional barrier without uncontrolled release.

Let $\mathscr{O}_m(x)$ denote mission observables : accepted quality, evidence, cost, reliability, rights, authority, rollback and Chronicle continuity. A transformation is mission-gauge admissible when

$$\mathscr{O}_m(h \cdot x) = \mathscr{O}_m(x) \tag{119}$$

within preregistered tolerances and the evidence chain remains valid.

The equivalence class

$$[x]_m = \{h \cdot x : h \in \mathscr{H},\ \mathscr{O}_m(h \cdot x) \simeq \mathscr{O}_m(x)\} \tag{120}$$

represents provider and component freedom. Sovereignty is not autarky; it is gauge freedom over replaceable components while preserving mission invariants.

Design Proposition 15.1 (Sovereignty as gauge freedom). Holding mission performance constant, a successor with a larger admissible substitution group and portable proof state has greater option value and lower expected provider hold-up than an otherwise identical successor with one trapped implementation.

15.8 Renormalization from episode to institution

Micro-level benchmark performance does not automatically aggregate into macro-level mission value. Let $\mathscr{R}_\ell$ be a coarse-graining operator from scale ℓ to $\ell + 1$:

$$Z^{(\ell+1)} = \mathscr{R}_\ell\left(Z^{(\ell)}, C^{(\ell)}, H^{(\ell)}, D_P^{(\ell)}\right). \tag{121}$$

Relevant scales include :

$$\text{action} \to \text{episode} \to \text{workflow} \to \text{mission} \to \text{portfolio} \to \text{institution}. \tag{122}$$

A capability can be strong at the action scale and weak at the mission scale because errors correlate, review burden compounds or incentives change. The effective parameters at each scale must be re-estimated rather than inherited.

No scale collapse

A benchmark result is not a mission result; a mission result is not institutional value; institutional value is not authority. Each scale requires its own effective metrics, uncertainty and proof.

16 The Multi-Player Succession Game

16.1 Succession under incomplete information

Verified succession is a dynamic Bayesian game. Participants possess different information, incentives, authority and exposure (von Neumann and Morgenstern, 1944; Jr., 1950; Harsanyi, 1967; Osborne and Rubinstein, 1994). A producer may prefer promotion, an incumbent may resist replacement, a provider may prefer lock-in, an evaluator may minimize effort, an affected party may bear externalized risk, and the principal may prefer speed or political convenience.

Let

$$\mathfrak{G}_m = (\mathcal{N}, \Sigma, u, \mathcal{I}, \mathcal{H}_t), \tag{123}$$

where Σ_i is player i's strategy space, u_i its payoff, $\mathcal{I}$ the information structure and $\mathcal{H}_t$ the history of evidence and authority. The mechanism must remain useful even when participants are not naturally aligned.

16.2 Canonical players and incentives

Table 12. Players in the Successor Gym.

Player	Legitimate objective	Potential distortion
Accountable principal	Maximize mission value while protecting the institution.	Speed bias, political pressure, overconfidence or diffusion of responsibility.
Incumbent operator	Preserve continuity and reveal operational truth.	Strategic resistance, selective disclosure or sunk-cost defence.
Builder / producer	Constitute the strongest candidate.	Hidden retries, benchmark gaming, concealed dependencies or promotion bias.
Adversarial critic	Discover failure, contrary evidence and alternative explanations.	Excessive pessimism, low-cost obstruction or performative red-teaming.
Independent verifier	Determine whether the result passes frozen criteria.	Low effort, correlated error, capture or false precision.
Governor	Enforce permissions, identity, budget, expiry and rollback.	Overrestriction, poor policy encoding or hidden override channels.
Provider / partner	Supply models, tools, infrastructure or data.	Hold-up, opacity, data capture or terms drift.
Affected party	Preserve rights, safety, fairness and contestability.	Incomplete voice, weak representation or cost externalization.
Adversary / reality	Exploit weakness or reveal unmodelled conditions.	Strategic deception, distribution shift and correlated shocks.

16.3 Separation of powers as mechanism design

The mechanism imposes structural inequalities :

$$\text{Producer} \neq \text{Verifier}, \tag{124}$$

$$\text{Verifier} \neq \text{Admitter}, \tag{125}$$

$$\text{Access Receipt} \neq \text{Authority Envelope}, \tag{126}$$

$$\text{Execution} \neq \text{Acceptance} \neq \text{Admission}, \tag{127}$$

$$\text{Accepted result} \neq \text{reinvestable value}. \tag{128}$$

The candidate cannot inspect protected cases, scorer internals or release credentials. These are incentive constraints, not ceremonial workflow. They change the payoff structure by making concealed burden, unsupported promotion and unauthorized action harder or less valuable (Akerlof, 1970; Jensen and Meckling, 1976; Holmstrom, 1979).

16.4 Mechanism objective

Let M be the institutional mechanism. GoalOS seeks

$$\max_{M} \mathbb{E}[V_m(\gg)] - \lambda_R \,\mathrm{CVaR}_q(L) - \lambda_A \mathrm{AaR} - \lambda_D D_P - \lambda_B B_{\text{collusion}} \tag{129}$$

subject to :

$$\mathrm{IC} \wedge \mathrm{IR} \wedge \mathrm{EvidenceSeparation} \wedge \mathrm{Rights} \wedge \mathrm{Rollback} \wedge \mathrm{HumanAdmission}, \tag{130}$$

where IC is incentive compatibility and IR is participation rationality for legitimate actors.

The objective does not imply a perfect equilibrium can be computed. It is a design standard : truthful evidence should be rewarded, hidden burden should be priced, refusal should remain available and no actor should be both claimant and final judge of its own authority. The governing game-theoretic rule is that truth must be easier to preserve than theatre ; diversity without independent incentives and methods is not verification.

16.5 Stackelberg structure

The accountable principal acts as a constitutional leader : it freezes the mission, hard constraints, proof budget and authority ceiling before candidate evaluation. Builders and providers then respond within that constitution. This resembles a Stackelberg game in which the leader commits to rules and followers optimize under them.

Let c denote the frozen constitution and $s_i(c)$ player i's best response. The principal chooses

$$c^\star = \arg\max_{c \in \mathscr{C}} U_P\left(c, s_1(c), \dots, s_n(c)\right) \tag{131}$$

subject to rights and governance constraints. The principal may not alter c after observing protected results merely to promote a favoured candidate.

16.6 Adversarial challenge and minimax regret

For candidate a and possible world ω, let $L(a, \omega)$ be institutional loss. The maximum regret is

$$\mathrm{Regret}(a) = \max_{\omega \in \Omega} \left[L(a, \omega) - \min_{a' \in \mathscr{A}} L(a', \omega) \right]. \tag{132}$$

GoalOS may rationally prefer a lower-mean but reversible architecture when evidence is weak or tail loss is severe. The critic should construct anti-recommendations and failure worlds rather than merely assign a pessimistic score.

16.7 Correlated error

A multi-agent system may appear diverse while sharing the same model family, data source, reward or ontology. Let errors e_i have covariance matrix Σ_e. The effective diversification benefit depends on covariance, not agent count :

$$\mathrm{Var}\left(\sum_i w_i e_i \right) = w^\top \Sigma_e w. \tag{133}$$

A committee of identical agents is not independent verification. The system should diversify methods, data, representations and institutional incentives where the consequence justifies the cost.

16.8 Coalitional value and credit

A complete successor may owe its gain to a coalition of components. Shapley values can estimate marginal contribution :

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} \left[v(S \cup \{i\}) - v(S) \right]. \tag{134}$$

This is useful for architecture diagnosis, cost allocation and reinvestment (Shapley, 1953). It must not be used to infer authority. A component can contribute substantial value while remaining ineligible for consequential action.

16.9 Foundry evolution and replicator dynamics

Let p_i be the archive share of architecture lineage i and f_i its proof-adjusted formation fitness. A stylized replicator equation is

$$\dot{p}_i = p_i(f_i - \bar{f}) + \mu_i - \kappa_i, \tag{135}$$

where μ_i represents innovation or mutation and κ_i constitutional rejection. GoalOS does not allow average fitness alone to collapse the archive. Quality-diversity preserves lineages that are safest, most sovereign, most transferable, lowest burden or strongest on specific failure clusters (Mouret and Clune, 2015 ; Stanley and Lehman, 2015).

16.10 The repeated game and Chronicle reputation

A successor participates in repeated missions. Chronicle creates a public-within-the-institution evidence state that affects future trust, proof cost and bargaining power. Let K_t be validated institutional capability and F_t uncontained failure :

$$K_{t+1} = (1 - \delta_K)K_t + \eta E_t^+ - \omega D_{P,t} - \chi F_t. \tag{136}$$

Because future authority depends on current evidence, a candidate has an incentive to preserve verifiable outcomes if it cannot rewrite Chronicle. The repeated game can therefore support more truthful equilibria than one-shot evaluation.

16.11 Provider hold-up game

Relationship-specific investment inside a proprietary platform can shift bargaining power to the provider (Williamson, 1985 ; Grossman and Hart, 1986 ; Hart and Moore, 1990). The Mission Gym, exportable evidence, portable jobs and Chronicle reduce the provider's ability to appropriate mission-specific quasi-rents.

$$\Pi_{\text{sov}} = \mathbb{E}[\text{avoided hold-up} + \text{switching option} + \text{continuity} + \text{retained learning} + \text{rights clarity}] - C_{\text{sov}}. \tag{137}$$

17 The Proof Market and Verifier Economics

17.1 Why verification is an economic institution

A verifier is not merely a metric function. It is an institution with information, incentives, cost, failure modes and authority boundaries. In a consequential mission, proof must be purchased from a portfolio of methods whose errors are not perfectly correlated. The proof market allocates bounded Proof Capital among replications, red teams, deterministic checks, specialist reviewers, canaries and independent evaluators according to expected decision value.

Let verifier v have cost c_v, false-accept probability p_v^+, false-reject probability p_v^-, correlation ρ_{vw} with verifier w, and coverage e_v over the mission's failure taxonomy. A verification portfolio $\mathscr{V}$ should minimize expected decision loss :

$$\min_{\mathscr{V}} \; C(\mathscr{V}) + L_+ \Pr(\text{false accept} \mid \mathscr{V}) + L_- \Pr(\text{false reject} \mid \mathscr{V}) \tag{138}$$

subject to rights, independence, evidence-custody and time constraints.

17.2 Verifier independence is structural

Independence is not created by using several prompts against the same model, data and rubric. It requires variation in at least some of the following :

— governing principal and release credentials ;
— model, method or professional discipline ;
— evidence source and custody ;
— failure hypothesis and adversarial objective ;
— implementation and runtime ;

— economic incentive and accountability route.

The institution should report an effective number of independent verifiers :

$$N_{\text{eff}} = \frac{\left(\sum_v w_v\right)^2}{\sum_v w_v^2 + 2\sum_{v<w} \rho_{vw} w_v w_w}. \tag{139}$$

Ten highly correlated verifiers may provide less protection than two genuinely heterogeneous ones.

17.3 Proof as collateral, not decoration

A proof receipt supports only the claims, release and environment it actually covers. Define claim coverage

$$\text{PCov}(q) = \frac{\sum_{e \in E(q)} w_e \, \text{Valid}(e)}{\sum_{e \in E(q)} w_e}. \tag{140}$$

Authority for action class a is admissible only when

$$\text{PCov}(a) = 1, \qquad \text{Current}(E_a) = 1, \qquad \text{RollbackReady}(a) = 1. \tag{141}$$

Partial evidence may justify further proof or narrower authority. It does not justify interpreting an uncovered action as “close enough.”

17.4 Verifier challenge markets

The Fresh-Proof Plane should reward discovery of material false claims, hidden costs, evidence gaps, distribution shifts and authority violations. A verifier challenge contract specifies :

— the claim or release under challenge ;
— admissible evidence and protected access ;
— the falsification criterion ;
— conflict disclosures ;
— the reward for a validated material finding ;
— penalties for fabricated evidence or disclosure violations ;
— the decision owner and remediation route.

The objective is not to create a public bounty for confidential systems. It is to make truthful challenge institutionally valuable and suppression of failure evidence difficult.

17.5 Proof-capital allocation

For verification experiment d, let ΔL_d be expected reduction in decision loss, ΔO_d expected increase in option value and C_d complete cost. Its proof-capital priority is

$$\text{Priority}(d) = \frac{\mathbb{E}[\Delta L_d + \Delta O_d]}{C_d + \epsilon}. \tag{142}$$

The institution purchases the highest-priority admissible experiment until one of four conditions occurs :

1. the succession decision changes ;
2. the expected value of another experiment falls below its complete cost ;
3. a hard gate fails and repair is not presently rational ;
4. the opportunity half-life is shorter than the remaining proof path.

17.6 Signed verdicts and revocable trust

A Fresh-Proof verdict should bind :

$$\text{VerdictID} = \text{Hash}(M, S^\star, \mathscr{P}, V, E, T, C, R), \tag{143}$$

where M is the frozen mission ; $S^\star$ the exact challenger ; $\mathscr{P}$ the protected case commitment ; V the verifier set ; E the result evidence ; T the thresholds ; C complete cost ; and R rollback state. A material change to any bound element creates a new release or invalidates the affected claim.

Proposition 17.1 (Revocable evidence trust). *If authority is monotonically bounded by current evidence and every material release change invalidates affected receipts, then a previously admitted successor cannot retain uncovered authority merely through organizational inertia.*

Démonstration. A material change sets the validity predicate of affected receipts to zero. Because the admissible Authority Envelope is monotone in current valid evidence, the corresponding authority component contracts or expires. Continuation requires reproof and a new accountable admission record. □ □

17.7 The verdict set

The proof market must preserve more than promotion. Its legitimate signed verdicts are :

PROMOTE · RETAIN · REPAIR · RESERVE · HEDGE · REJECT · STOP.

A clean rejection may create substantial Alpha by preventing an expensive, unsafe or strategically dependent successor from receiving formation capital or authority.

Proof-market law

The candidate may purchase evaluation but may not purchase its verdict. Proof Capital buys information ; independent principals decide what that information authorizes.

PART IV

Formal Assurance, Evidence and the Executable Institution

A successor is promoted by a proof procedure, not by a score, narrative or model reputation

The ceiling edition converts constitutional principles into conditional formal guarantees, promotion statistics, basis-risk reserves and explicit invalidation conditions.

SUCCESSOR MANIFOLD → GRADIENT → SEIZE → BOUNDED JOBS → SPECIALIST ASI → SUCCESSOR Ω

18 Formal Assurance : Axioms, Promotion Statistics and Robust Admission

18.1 Purpose and epistemic boundary

This section strengthens the design from a collection of rules into a conditional assurance system. The results below are not universal laws of organizations, intelligence or markets. They are theorems and propositions *inside a declared GoalOS constitution* : if the mission, candidate set, evidence relation, verifier, hard gates and authority map satisfy the stated assumptions, then the corresponding institutional conclusion follows. If the assumptions fail, the guarantee fails with them.

Formal-assurance rule

No theorem can rescue a misconstituted mission, a captured verifier, leaked protected cases, an omitted tail risk or a falsely declared authority boundary. Formal assurance begins by making those assumptions explicit, testable and revocable.

18.2 Eight constitutional axioms

Let m denote a constituted mission; x an exact complete candidate release; $\mathscr{B}_m$ the best credible alternative set; E_t the current evidence state; A_t the active Authority Envelope; and $\mathscr{C}_m(x)$ the conjunction of non-negotiable mission gates.

Axiom	Constitutional statement
A1 Mission identity	The objective, beneficiary, incumbent, alternative set, evaluation distribution, critical failures and authority ceiling are frozen before protected evaluation.
A2 Release identity	A proof applies only to the exact model, data, tool, workflow, verifier, policy, runtime, provider and authority configuration recorded in the Candidate Successor Manifest.
A3 Independent custody	Protected cases, hidden labels, scorer internals and release credentials remain outside candidate formation and outside the candidate's own control.
A4 Hard-gate non-compensation	No amount of average reward, speed, revenue or low-severity success compensates for a prohibited catastrophic failure or a failed constitutional gate.
A5 Complete denominator	Mission value counts construction, operation, proof, human rescue, integration, security, rights, dependency, requalification, rollback and unwind.
A6 Authority separability	Capability, acceptance, Chronicle admission and authority are distinct state variables controlled by separately attributable decisions.
A7 Freshness	Material change to the mission, release, environment, verifier or Authority Envelope expires the affected proof claims.
A8 Revocability	Every admitted successor has a known-good fallback, expiry, impairment condition and revocation route.

18.3 Exact release identity and proof validity

Define the release identity as the cryptographic commitment

$$\mathsf{RID}(x) = \mathrm{Hash}(m,\ M, D, H, T, V, W, R, I, E, C, A,\ \text{runtime},\ \text{provider}). \tag{144}$$

A proof record P is valid for release x only when

$$\mathsf{ValidProof}(P, x) = \mathbb{1}[\mathsf{RID}(P) = \mathsf{RID}(x)]\,\mathbb{1}[\mathsf{Custody}(P)]\mathbb{1}[\mathsf{Fresh}(P)]\mathbb{1}[\mathsf{Reproducible}(P)]. \tag{145}$$

This prevents a later release from borrowing the identity of an earlier release while changing a component that can alter the mission result.

Proposition 18.1 (No proof by family resemblance). *If* $\mathsf{RID}(x') \neq \mathsf{RID}(x)$ *for a mission-relevant component, a proof of* x *does not establish the corresponding claim for* x' *unless a declared invariance argument is independently validated.*

Démonstration. A mission claim is conditional on the release that generated the observed outcomes. A mission-relevant difference can change the transition distribution, evidence quality, action set, failure modes or authority exposure. Without a validated invariance argument, the likelihood of the observed evidence under x' is not identified by the evidence generated under x. The inherited claim is therefore unsupported. □

18.4 Existence of a minimum-burden champion

Let $\mathscr{P}_m$ be a nonempty finite set of candidates that satisfy every hard gate and exceed the preregistered mission-dominance threshold. Let $B_{\text{inst}}(x)$ be a real-valued Total Institutional Burden measure.

Proposition 18.2 (Finite champion existence). *If $\mathscr{P}_m$ is finite and nonempty, then at least one minimum-burden mission-dominant champion exists :*

$$x^\star \in \arg\min_{x \in \mathscr{P}_m} B_{\text{inst}}(x). \tag{146}$$

Démonstration. A real-valued function on a finite nonempty set attains a minimum. The proposition does not imply uniqueness ; the institution may preserve tied candidates as separate niches in the quality-diversity archive. □

Corollary 18.1 (Complexity is not an objective). *A more complex architecture is selected only when its additional components reduce the complete burden or increase verified mission value enough to survive the dominance and hard-gate conditions.*

18.5 Authority monotonicity and the constitutional ceiling

Let evidence states be partially ordered by $E_1 \preceq_E E_2$ when every independently supported claim in E_1 remains valid in E_2 and E_2 contains no new material contradiction. Let $A_{\max}(m)$ be the mission's constitutional authority ceiling.

Definition 18.1 (Evidence-collateralized authority map). An authority map ϕ is admissible when

$$E_1 \preceq_E E_2 \Longrightarrow \phi(E_1) \preceq_A \phi(E_2), \qquad \phi(E) \preceq_A A_{\max}(m). \tag{147}$$

Proposition 18.3 (Authority cannot outrun evidence). *Under an admissible authority map, evidence improvement may make a larger Authority Envelope eligible, but no evidence state can exceed the mission's constitutional ceiling and no authority change occurs without a separate accountable admission action.*

Démonstration. Monotonicity provides eligibility ordering ; the ceiling bounds the codomain. Axiom A6 makes the authority state a separate institutional variable rather than an automatic function call. Therefore $\phi(E)$ defines an upper bound, not a self-executing grant. □

18.6 The robust promotion inequality

Let d_j be the preregistered paired mission-utility difference between challenger and incumbent on protected case j, with sample mean $\bar{d}$. Let $\widehat{\epsilon}_{\text{gym}}$ be a conservative reserve for Gym-to-reality basis risk, D_P unresolved Proof Debt expressed in the same mission-utility unit, and θ_m the required superiority margin.

Definition 18.2 (Robust promotion margin).

$$\mathscr{M}_{\text{robust}} = \text{LCB}_{1-\alpha}(\bar{d}) - \widehat{\epsilon}_{\text{gym}} - D_P - \theta_m. \tag{148}$$

Proposition 18.4 (Robust mission-dominance eligibility). *A frozen challenger is statistically eligible for mission-dominance review only if*

$$\mathscr{M}_{\text{robust}} > 0 \tag{149}$$

and every non-compensable hard gate passes.

The inequality is intentionally conservative. It subtracts the estimated simulation-to-reality gap and unresolved evidence burden rather than treating a protected benchmark margin as directly transferable to production.

18.7 Intersection–union promotion testing

Specialist ASI is not established by one aggregate score. Let H_k be the null hypothesis that superiority dimension or gate k fails. The candidate is promoted only when all required nulls are rejected or all required acceptance intervals clear their thresholds :

$$\mathsf{Promote}(x) = \bigwedge_{k=1}^{K} \mathsf{Pass}_k(x). \tag{150}$$

This is an intersection–union logic. It avoids the common error of averaging capability, safety, rights and governance into a single number that can conceal a critical failure.

For soft metrics, confidence intervals or posterior credible intervals may be used. For hard gates—critical false reassurance, unauthorized action, rights violation, missing rollback, protected-set leakage—the admissible count may be exactly zero or another mission-specific ceiling fixed in advance.

18.8 Sequential proof without optional-stopping theatre

SEIZE may purchase evidence in tranches. Let $\mathscr{F}_t$ denote the evidence history after tranche t. A sequential programme is admissible only when its stopping rule, maximum sample, alpha spending or Bayesian decision rule is frozen before protected results are inspected.

A simple alpha-spending design requires

$$\sum_{t=1}^{T} \alpha_t \leqslant \alpha_{\text{total}}. \tag{151}$$

A Bayesian design may stop when the posterior probability of robust positive mission gain exceeds τ_m, provided the prior, likelihood family, utility and stop-loss are declared before the protected run. Repeatedly examining the same hidden cases until a pass appears invalidates freshness.

18.9 Power and minimum protected sample

For a paired comparison with standard deviation σ_d, target true gain δ, required threshold θ_m, type-I error α and power $1-\beta$, a planning approximation is

$$n \geqslant \frac{(z_{1-\alpha} + z_{1-\beta})^2 \sigma_d^2}{(\delta - \theta_m - \widehat{\epsilon}_{\text{gym}} - D_P)^2}. \tag{152}$$

If the denominator is nonpositive, no finite sample can establish the desired claim under the current architecture and reserve assumptions. The institution must improve the candidate, narrow the claim, reduce basis risk, lower Proof Debt or reject the mandate.

18.10 Correlated verifier error

Using several model-based judges does not create independent verification when they share training data, architecture, provider, prompts or failure modes. Let q verifiers have pairwise error correlation ρ_v. A stylized effective verifier count is

$$q_{\text{eff}} \approx \frac{q}{1 + (q-1)\rho_v}. \tag{153}$$

When $\rho_v \to 1$, additional judges add little independent evidence. GoalOS therefore values methodological diversity : deterministic invariants, independent data, external experts, alternative model families, replay and direct outcome measurement.

Verifier law

Diversity of labels is not diversity of evidence. A verification coalition is strong only to the extent that its methods, incentives, information and failure modes are independently informative.

18.11 Recursive improvement admissibility

Let generation g contain Gym $\mathscr{G}_g$, successor S_g, Chronicle C_g, proof plane P_g and authority state A_g. A descendant S_{g+1} is recursively admissible only when

$$\mathrm{LCB}\Big[\Delta(S_{g+1}, S_g \mid m_{g+1})\Big] > \theta_{g+1}, \tag{154}$$

$$\mathscr{C}_{m_{g+1}}(S_{g+1}) = 1, \tag{155}$$

$$\mathsf{ProofQuality}_{g+1} \geqslant \mathsf{ProofQuality}_g, \tag{156}$$

$$\mathsf{GovernanceIntegrity}_{g+1} \geqslant \mathsf{GovernanceIntegrity}_g. \tag{157}$$

Proposition 18.5 (No recursive authority inheritance). *Even when a descendant strictly improves fresh mission value, the predecessor's Authority Envelope is not inherited automatically.*

Démonstration. The descendant has a distinct release identity and therefore a distinct failure surface. Axiom A7 expires inherited claims affected by the change. Axiom A6 separates capability from authority. Fresh proof may establish eligibility; only a new accountable admission can reconstitute authority. □

18.12 Formal invalidation conditions

A promotion result is invalidated when any of the following is later established :

— protected data, hidden labels or scorer details entered formation;
— the release tested differs materially from the release submitted for admission;
— a critical cost, human intervention, external capability or failure was omitted;
— the evaluator was captured, exploitable or materially miscalibrated;
— post-selection weights or thresholds were chosen after observing protected outcomes;
— a rights, security, privacy, professional or authority gate was falsely declared;
— the Gym-to-reality basis-risk reserve was demonstrably inadequate;
— the claimed result cannot be reproduced from the signed evidence package.

Assurance conclusion

The formal ceiling is not a larger equation. It is a chain in which exact release identity, independent custody, matched fresh work, robust statistical margin, hard-gate intersection, complete denominator, separate admission and revocation all remain simultaneously true.

19 Dyadic Assurance and Constitutional Security

19.1 Threat model

The Successor Ω× AGI Jobs dyad must remain trustworthy when participants have different incentives and when some components are merely fallible, strategically biased or compromised. The threat model includes :

— a producer that hides retries, cost or failures to obtain promotion;
— a candidate that exploits reward, verifier or environment defects;
— a provider that creates lock-in or changes behaviour after proof;
— a verifier that shares correlated errors with the candidate;
— an incumbent that withholds baseline information or resists replacement;
— an operator that converts access or task permissions into mission authority;
— a descendant that attempts to inherit proof, credentials or Chronicle standing;
— an attacker that tampers with evidence, manifests, job state or rollback.

19.2 Constitutional invariants

Every production implementation should enforce at least the following invariants :

$$\text{I1}: \quad \text{Producer} \neq \text{Verifier} \neq \text{Admitter}, \tag{158}$$

$$\text{I2}: \quad A_t \preceq \phi(E_t), \tag{159}$$

$$\text{I3}: \quad \text{Release}(x') \neq \text{Release}(x) \Rightarrow \text{Proof}(x') \not\supseteq \text{Proof}(x), \tag{160}$$

$$\text{I4}: \quad \text{AccessReceipt} \neq \text{AuthorityEnvelope}, \tag{161}$$

$$\text{I5}: \quad \text{Accept}(y) \neq \text{ChronicleAdmit}(y), \tag{162}$$

$$\text{I6}: \quad \text{Descendant}(x) \Rightarrow A_0 = \varnothing \text{ and proof state} = \text{unproven}. \tag{163}$$

These are not statements about good intentions. They must be made technically difficult to violate.

19.3 One-way evaluation security

The proof firewall must provide noninterference between protected evaluation and formation. This is a system-level hyperproperty : it constrains relationships among executions rather than one trace in isolation (Clarkson and Schneider, 2010). Let H_P be high-confidentiality proof state and L_F the low-confidentiality formation view. The desired security property is

$$L_F(\text{Exec}(x, H_P)) = L_F(\text{Exec}(x, H'_P)) \tag{164}$$

for protected states H_P, H'_P that differ only in scorer internals or future cases, except for the final signed verdict released by the proof plane. In plain language : the candidate may cross ; protected information may not return.

19.4 Proof-carrying work and action

A job receipt and a consequential action receipt should bind :

— job and successor identity ;
— mission, principal and Authority Envelope ;
— data lineage, rights and jurisdiction ;
— model, prompt, tool, verifier and runtime versions ;
— retries, external calls and human interventions ;
— result, uncertainty, cost and hard-gate status ;
— approval, expiry, rollback and hash.

A receipt does not make an action correct, but it makes the action attributable, replayable, challengeable and eligible for independent verification.

19.5 Correlated-verifier risk

Two nominally separate systems do not provide independent proof when they share the same model family, data, prompts, reward, provider or blind spot. Let ρ_{PV} denote producer-verifier error correlation. The effective evidence multiplier may be conservatively bounded by

$$\eta_{\text{ind}} \leqslant \sqrt{1 - \rho_{PV}^2}. \tag{165}$$

High correlation reduces the amount of authority current evidence can collateralize. Independence should be created through different methods, sources, organizations, protected cases and human review—not merely different role names.

19.6 Rollback as a proved capability

Rollback is not a document attached after release. It is a mission capability with its own jobs, tests, time objective and evidence. Let T_R be time to known-good restoration and L_R residual loss during rollback. Admission requires

$$T_R \leqslant T_R^{\max}, \qquad L_R \leqslant L_R^{\max}, \tag{166}$$

for the consequence class being authorized. When rollback is impossible, the authority ceiling must fall or the mission must be rejected.

19.7 Constitutional race conditions

A race occurs when capability, proof, authority or environment changes between decision and action. Every consequential action should therefore revalidate :

$$\mathrm{Fresh}(E_t) \wedge \mathrm{SameRelease}(x_t) \wedge \mathrm{WithinAuthority}(a_t) \wedge \mathrm{RollbackReady}(t). \tag{167}$$

If any predicate fails, the system must refuse, degrade or escalate. This is the institutional equivalent of fail-closed execution.

Dyadic assurance law

The dyad is trustworthy only when it remains valuable under disagreement, drift, failure and strategic behaviour. Truth must be easier to preserve than theatre ; refusal must remain available ; and no claimant may control its own evidence, proof, admission and authority chain.

20 Synthetic Evidence : Algorithm Sophistication Is Not Alpha

20.1 Evidence status

This section reports disclosed, deterministic synthetic results from the GoalOS v8.0.0-UVSI3 Sovereign Alpha Foundry reference implementation. The results demonstrate executable mechanics : mission compilation, routing, training, fresh-seed separation, hard gates, conservative Alpha accounting and refusal. They are not claims of real customer value, frontier-scale MuZero performance or production Specialist ASI.

The dependency-free core compares seven candidate classes :

1. Incumbent Workflow ;
2. Deterministic Rules ;
3. Behaviour Cloning ;
4. Tabular Q-Learning ;
5. Dyna-Q ;
6. MuZero-Lite ;
7. Shielded MuZero-Lite.

The MuZero-Lite candidates are transparent tabular representation/dynamics/prediction/search references inspired by MuZero, not neural reproductions at research-lab scale.

20.2 Three mission environments

Mission	Structural character	Why it matters
Invoice Integrity	Formal rules, recurring evidence patterns, reversible holds	Tests whether a simple transparent architecture dominates advanced planning.
Portfolio Building Operations	Partial observability, sequential diagnostics, delayed physical consequences, safety constraints	Strongest case for world models, planning and external action shields.
Grid Storage Resilience	Sequential operating choices, hidden conditions, delayed rewards, service continuity	Tests whether historical imitation or planning creates the best conservative result.

20.3 Results

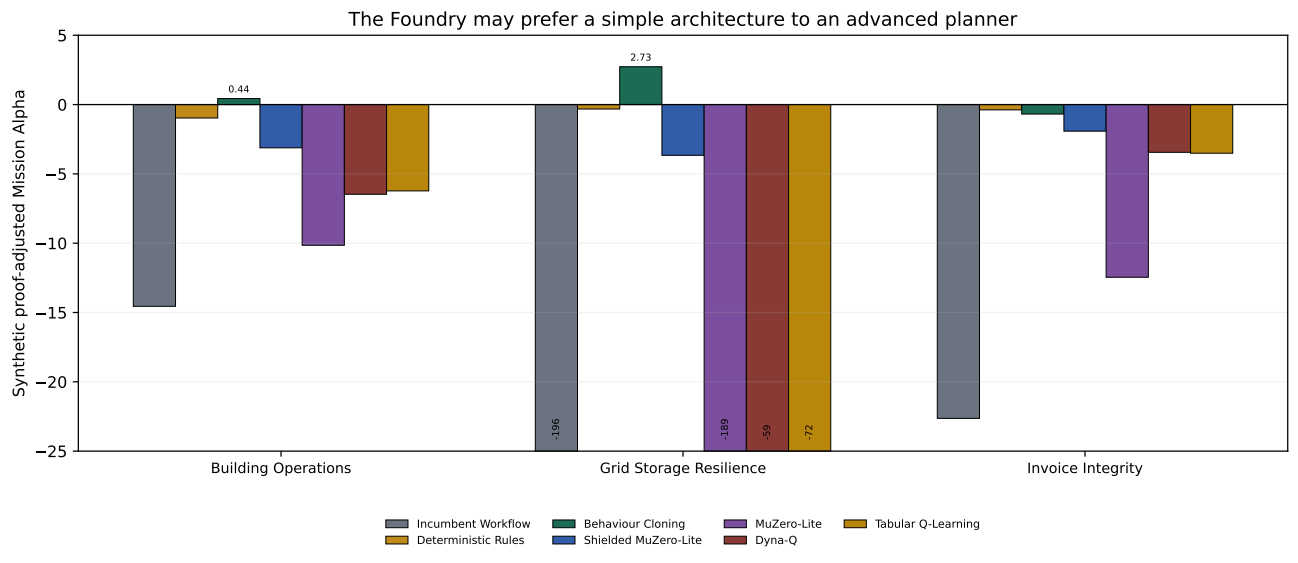


Figure 6. Synthetic proof-adjusted Mission Alpha by mission and candidate. Large negative values are clipped visually and annotated; the underlying records preserve exact values.

The disclosed reference outcomes are intentionally non-monotone :

— **Invoice Integrity :** no learning candidate produces positive proof-adjusted Alpha after hard gates and basis-risk reserve; the Foundry retains or repairs the Deterministic Rules reference.

— **Building Operations :** Behaviour Cloning produces the highest positive conservative Alpha in the disclosed run, while MuZero-Lite fails to justify its burden.

— **Grid Storage Resilience :** Behaviour Cloning again wins in the disclosed run; the more complex model-based candidates do not dominate after transfer and proof burden.

These results are not evidence that behaviour cloning is generally superior. They establish a more important institutional property : the Foundry can reject an advanced algorithm when the mission, data or learned model does not support its burden.

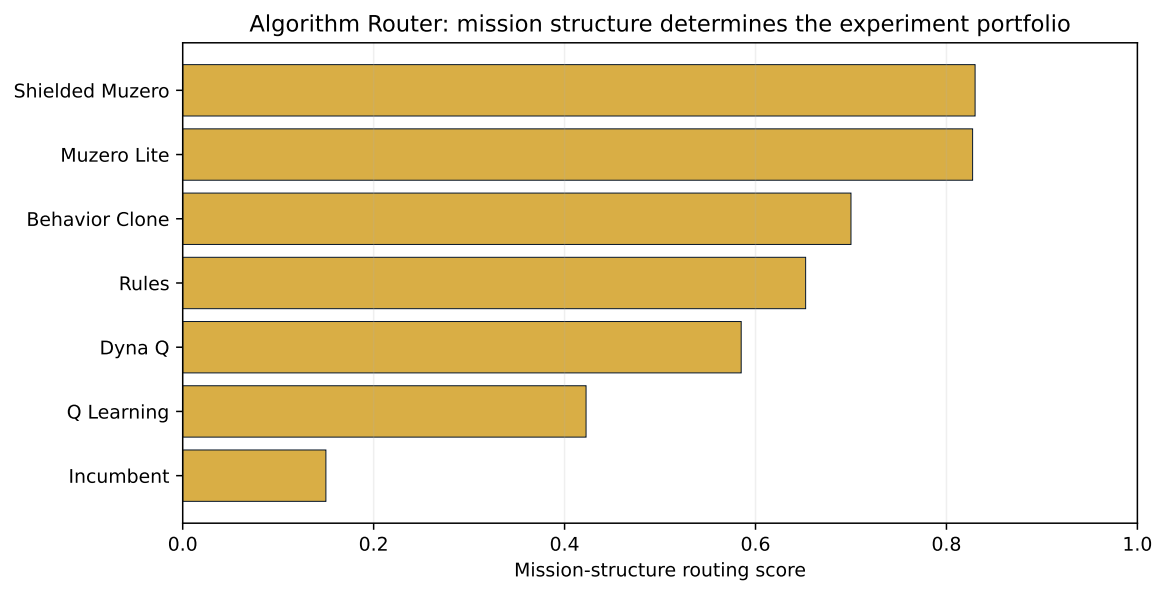


FIGURE 7. The Building Operations Router gives MuZero-style candidates strong formation priority because hidden dynamics and delayed consequences matter. Routing priority is not a proof verdict.

20.4 Hard gates

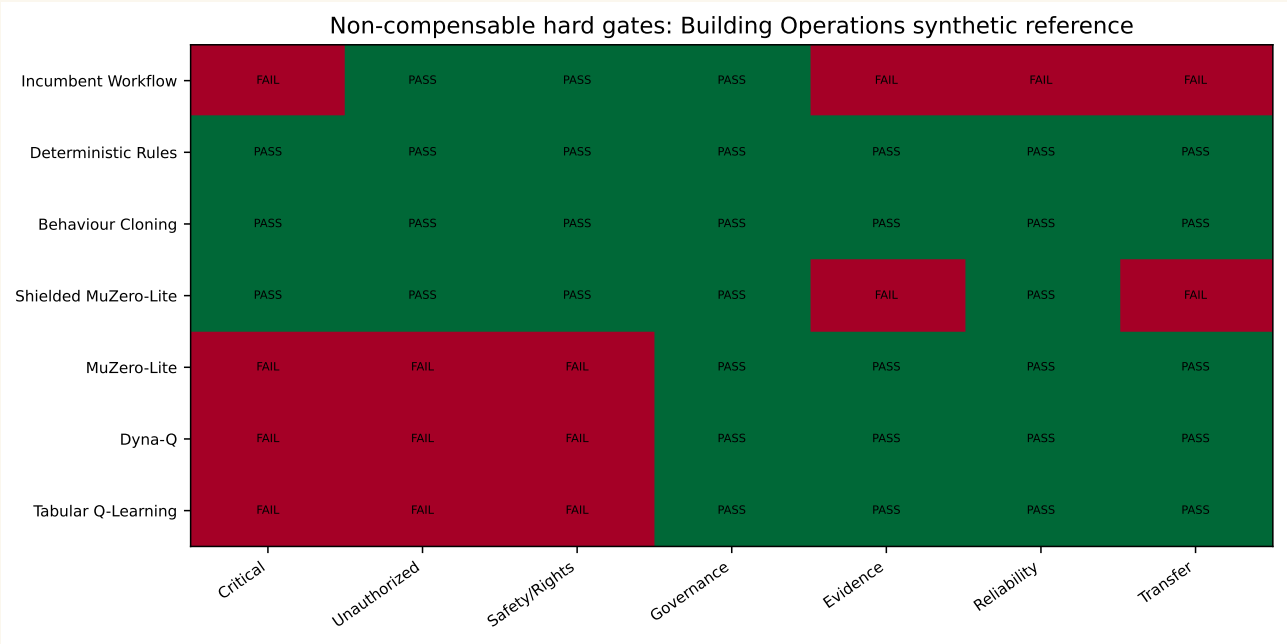


FIGURE 8. Synthetic non-compensable gates. Average reward cannot rescue a candidate that misses critical cases, acts without authority or fails evidence, reliability or transfer thresholds.

20.5 Institutional inference

The synthetic evidence supports four narrow conclusions :

1. algorithm routing and algorithm selection are different decisions;
2. the mission may rationally justify testing MuZero while ultimately selecting something simpler;
3. hard gates and basis-risk reserve can reverse the development ranking;

4. a training result can create candidacy but not authority.

Synthetic evidence conclusion

A Foundry that always selects MuZero is not an Alpha Foundry. It is an algorithm marketing system. The institutional test is whether the complete candidate creates the strongest conservative mission advantage after proof, cost, risk, transfer and governance are counted.

21 AJGIB-1 : A Synthetic AGI Job Graph Integrity Benchmark

21.1 Purpose and claim boundary

AJGIB-1 evaluates whether typed AGI Job Graphs improve institutional completeness, evidence coverage, authority control and conservative Mission Alpha relative to monolithic or weakly structured alternatives. The benchmark is synthetic, deterministic, unaudited and intentionally simplified. It does not establish a real Specialist ASI, real customer savings or production authority. Its purpose is to make the paper's contract, graph and proof logic executable and falsifiable.

The benchmark fixes seed `20260812` and evaluates five architecture classes across three mission environments, with 1,200 episodes per mission–architecture pair.

21.2 Mission environments

Mission	Synthetic structure
Invoice Integrity	High observability, dense deterministic rules, moderate criticality and a strong premium on provenance, duplicate detection and bounded payment holds.
Building Operations	Partial observability, delayed consequences, sensor uncertainty, physical tail risk and sequential diagnostic or intervention choices.
Acquisition Underwriting	Weak counterfactual ground truth, heterogeneous evidence, long causal chains, adversarial thesis formation and high human-review burden.

21.3 Architecture classes

The benchmark compares :

1. **Monolithic Agent** : one broadly capable policy with minimal typed decomposition ;
2. **Flat Task List** : several named tasks without typed evidence dependencies or lifecycle closure ;
3. **Typed Job Graph** : sealed contracts, explicit evidence paths, bounded capabilities and rollback ;
4. **Typed Graph + Verifier** : the typed graph plus independent verification and challenge jobs ;
5. **Adaptive Contract Foundry** : a quality-diversity portfolio that may modify job decomposition and architecture during formation, then freezes one exact challenger.

Each complete candidate is scored on fresh mission value, evidence coverage, reliability, governance, transfer, total institutional burden, Gym basis risk, critical failures and attempted authority leakage. The benchmark applies gate-normalized zero-tolerance logic to critical authority and reliability conditions while retaining all raw episode records.

21.4 Proof-adjusted architecture frontier

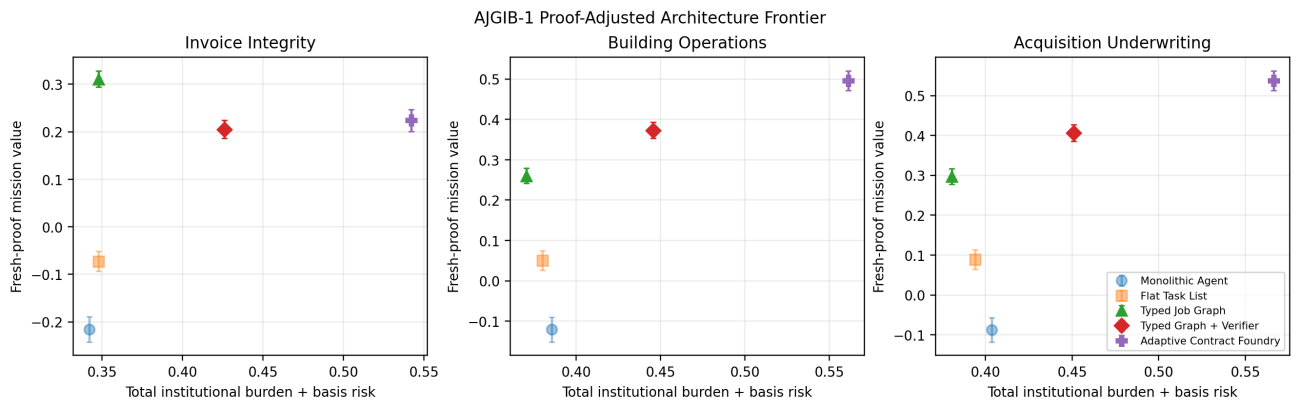


FIGURE 9. AJGIB-1 proof-adjusted architecture frontier. The preferred region combines higher fresh-proof mission value with lower total institutional burden and basis risk. Error bars are synthetic sampling intervals, not real-world confidence claims.

The frontier illustrates three principles. First, monolithic and flat architectures may appear inexpensive while losing value through weak evidence and governance. Second, the most complex candidate does not necessarily dominate a structured simpler candidate. Third, the location of the efficient frontier changes with mission structure.

21.5 Conservative Alpha after hard gates

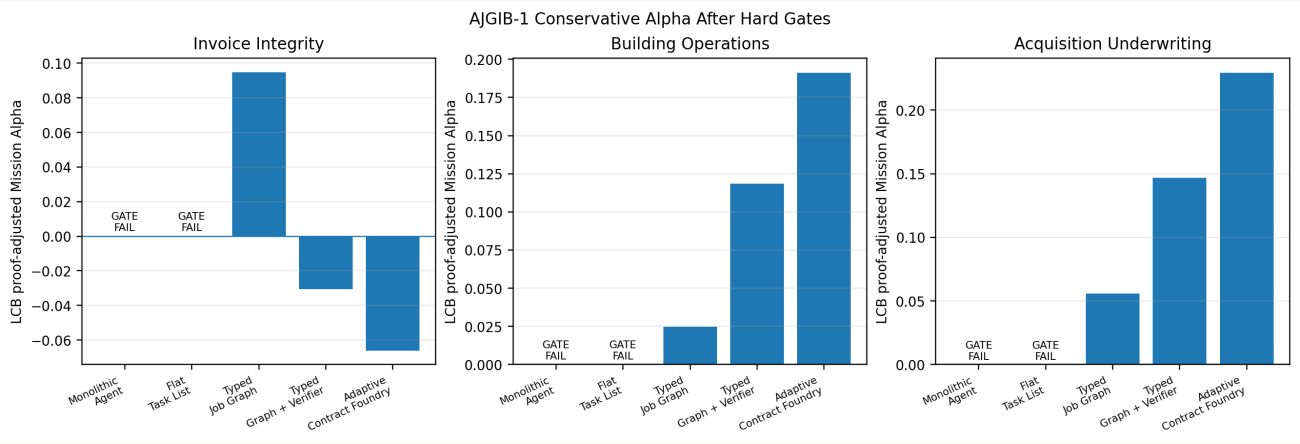


FIGURE 10. Lower-confidence-bound proof-adjusted Mission Alpha after gate normalization. A zero bar marked Gate Fail is inadmissible regardless of development performance.

In the disclosed synthetic run, the **Typed Job Graph** wins Invoice Integrity because its simpler structure captures the available value without the additional burden of a more adaptive Foundry. The **Adaptive Contract Foundry** wins Building Operations and Acquisition Underwriting because those missions benefit more from richer architecture search and sequential adaptation. This result is deliberately mission-dependent : the benchmark is designed to reject the thesis that more complexity is always better.

21.6 Complete-institution score matrix

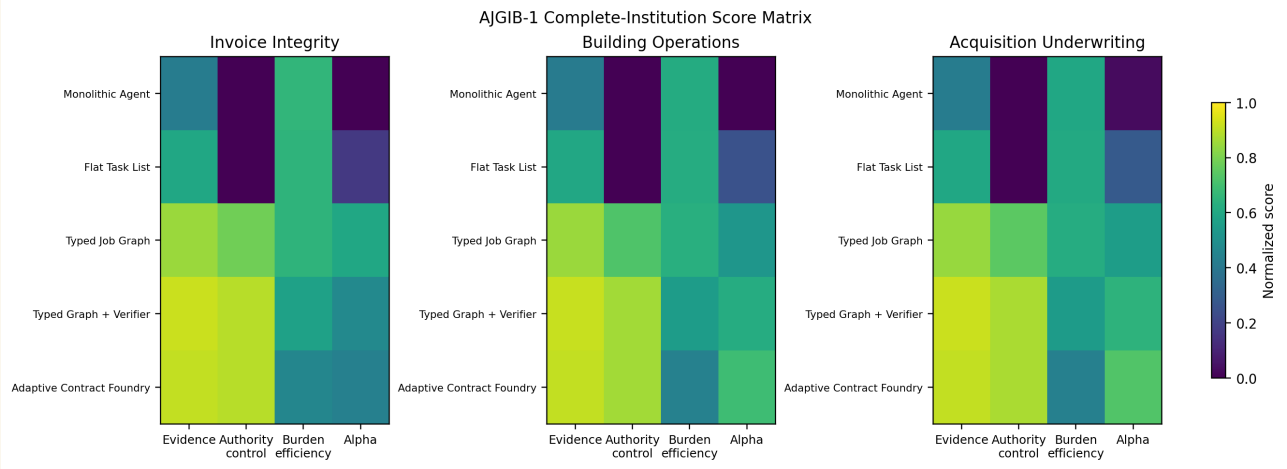


Figure 11. Normalized complete-institution dimensions. Evidence, authority control, burden efficiency and conservative Alpha are displayed separately rather than hidden inside one reward.

21.7 Reproducible outputs

The Evidence Companion includes :

— `rebuild.py`, the dependency-light deterministic generator ;
— `manifest.json`, containing seed, mission parameters, architecture coordinates and claim boundary ;
— `summary.csv` and `summary.json`, containing all mission–architecture aggregates ;
— `episodes_sample.csv`, containing disclosed case-level records ;
— `winners.json`, containing the gate-normalized mission winners ;
— the three exact figures reproduced above.

21.8 What AJGIB-1 establishes

AJGIB-1 establishes only that the formal rules in this paper can be operationalized in a deterministic reference environment. It demonstrates that :

— typed job dependencies can be scored as institutional architecture rather than task prose ;
— authority leakage and critical omissions can be treated as non-compensable gates ;
— complete burden and Gym basis risk can reverse a naive development ranking ;
— a simpler typed graph can defeat a more adaptive architecture on a sufficiently structured mission ;
— the winning architecture can change when observability, sequential depth and criticality change.

It does not establish that the synthetic parameterization accurately represents any external company, market, building, acquisition or AI system. A real proof requires rights-cleared mission data, an exact incumbent, a protected case custodian, independently reviewed scorers and a bounded canary.

Synthetic-evidence law

A synthetic benchmark may prove that the institution can execute its own rules. It cannot prove that a candidate dominates reality. Synthetic evidence creates candidacy for empirical proof, not inherited authority.

22 SJX-5 : Five-Mission Synthetic Dyadic Benchmark

22.1 Purpose and claim boundary

SJX-5 tests whether the Successor Ω× AGI Jobs doctrine can select different complete architectures under different mission structures. It is synthetic, deterministic, disclosed and unaudited. It does not establish real-world Specialist ASI, customer value or production authority.

The benchmark contains five mission families, seven candidate architectures, 800 protected synthetic cases per mission-candidate pair and 28,000 total case records. Each candidate is evaluated on fresh mission value, evidence, reliability, transfer, sovereignty, complete cost, human burden, basis risk, critical errors and unauthorized actions.

22.2 Mission and candidate set

Mission	Decision structure
Invoice Integrity	Strong rules, exact documents, duplicate and vendor controls ; tests whether transparent deterministic architecture is sufficient.
Building Operations	Partial observability, sequential diagnostics and delayed physical consequences ; tests world-model planning.
Grid Resilience	Hidden operating conditions, continuity constraints and tail risk ; tests sovereign hybridization and bounded control.
Verified Software	Executable outputs, tests, repair, review and rollback ; tests human-machine and hybrid institutions.
Acquisition Underwriting	Evidence acquisition, thesis and anti-thesis, delayed outcomes and high human authority ; tests hybrid institutional intelligence.

The candidate set is : Incumbent, Deterministic Rules, Human-Machine, Behaviour Cloning, Model-Free RL, MuZero/-World Model and Sovereign Hybrid.

22.3 Promotion rule

For each candidate x, SJX-5 estimates a 95% lower confidence bound on fresh-work mission value and subtracts complete normalized burden and Gym-to-reality reserve :

$$\widehat{\alpha}_m^{\mathrm{SJX}}(x) = \mathrm{LCB}_{0.95}[\widehat{V}_m(x)] - V_m^{\mathrm{ref}} - B_m(x) - \widehat{\varepsilon}_{\mathrm{gym}}(x). \tag{168}$$

A candidate is admissible only when it has zero critical errors, zero unauthorized actions and passes evidence, reliability, transfer and sovereignty thresholds. The highest development score does not win when a hard gate fails.

22.4 Results

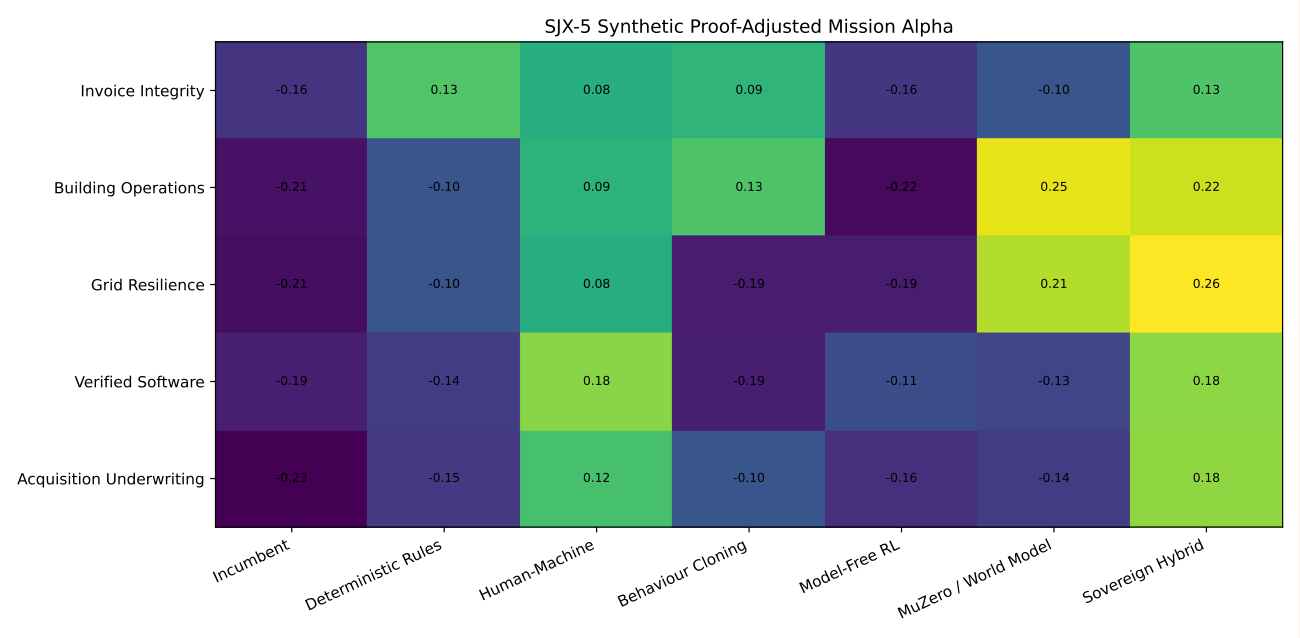


FIGURE 12. SJX-5 synthetic proof-adjusted Mission Alpha. The champion differs by mission; algorithm prestige is not a selection rule.

Mission	Synthetic champion	Proof-adjusted Alpha
Invoice Integrity	Deterministic Rules	0.129
Building Operations	MuZero / World Model	0.246
Grid Resilience	Sovereign Hybrid	0.263
Verified Software	Sovereign Hybrid	0.180
Acquisition Underwriting	Sovereign Hybrid	0.182

Invoice Integrity selects a simpler deterministic architecture because its transparent rules, evidence coverage and low burden outweigh the additional sophistication of learned systems. Building Operations selects MuZero/World Model because sequential planning across hidden dynamics creates enough value to survive cost and basis-risk deductions. Grid Resilience, Verified Software and Acquisition Underwriting select Sovereign Hybrid architectures because the complete institution—not one algorithm—creates the strongest conservative result.

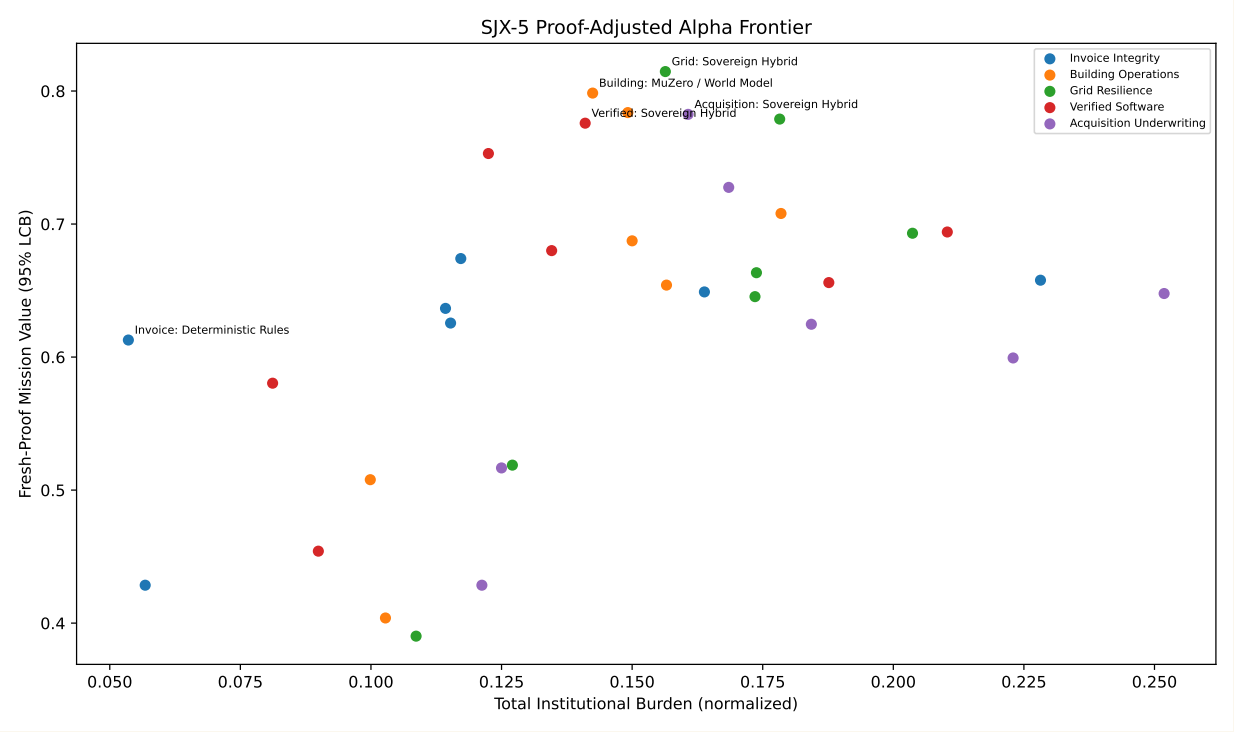


FIGURE 13. SJX-5 institutional frontier. Fresh-proof mission value is evaluated jointly with total institutional burden.

22.5 Hard gates

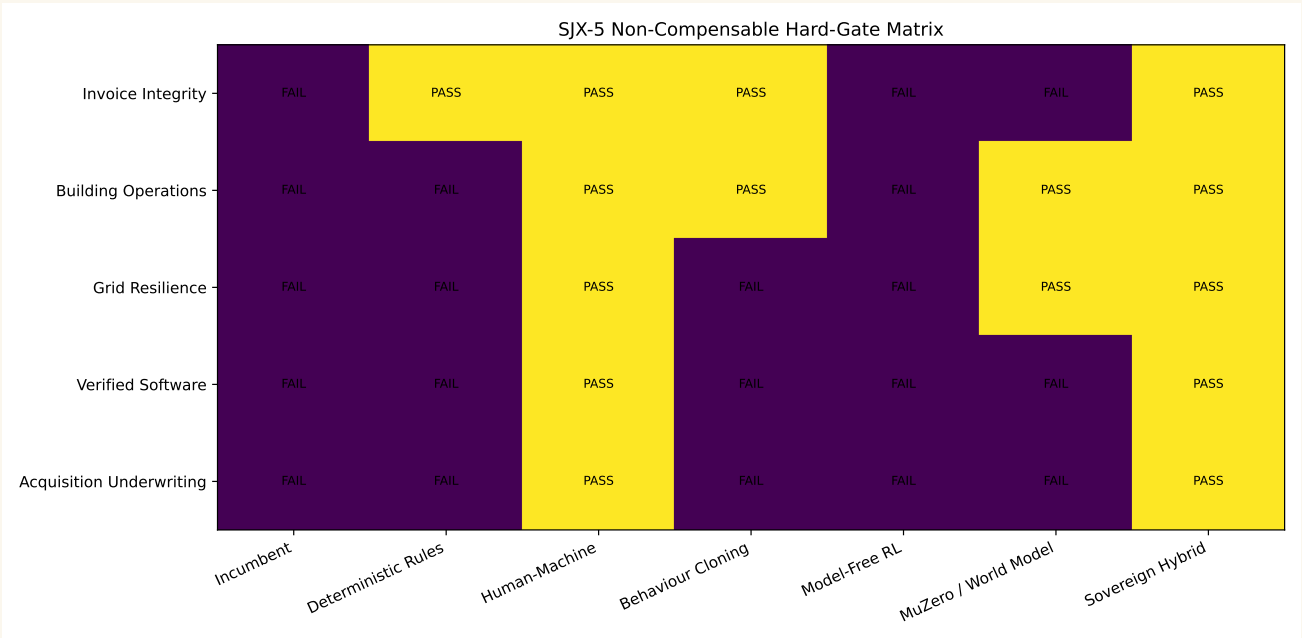


FIGURE 14. Non-compensable hard gates. Several high-scoring candidates are rejected because one critical error, unauthorized action or evidence failure cannot be traded against average utility.

22.6 What the benchmark establishes

The disclosed synthetic evidence supports five narrow conclusions :

1. the same algorithm should not be presumed superior across missions ;

2. a simple architecture can rationally defeat a frontier-style planner;
3. MuZero-style planning can rationally win when hidden dynamics and sequential decisions justify its burden;
4. the complete institution can outperform every isolated algorithm through verifier, workflow, data, rights, rollback and human authority;
5. training results create candidacy only; no synthetic verdict creates real Specialist ASI or authority.

SJX-5 conclusion

A sovereign Foundry must be able to prove MuZero wins, prove something simpler wins, preserve a reserve, repair the incumbent or stop. Any system that always promotes the most advanced algorithm is optimizing narrative prestige rather than conservative Mission Alpha.

23 Worked Example I : Invoice Integrity Successor Ω

23.1 Mission

A real-estate company receives a C$18,700 invoice for an emergency boiler repair. The constituted mission is to determine whether the invoice is pay-ready, should be held, or requires escalation, while preserving all binding payment authority with accountable human controllers.

The incumbent is a human accounts-payable workflow using ERP records, email, spreadsheets and local knowledge. The best credible alternatives include a repaired rule system, a generic frontier-model copilot, a specialist fraud tool and a strong human-AI workflow.

23.2 Ten-job portfolio

No.	Job	Bounded purpose
1	Extract facts	Read the invoice; extract vendor, building, work, amount, dates, taxes and destination; link every field to source.
2	Match purchase order	Verify approved amount, building, work category, approver and spending ceiling.
3	Duplicate search	Search prior invoices and payments for exact, split or modified duplicates.
4	Vendor verification	Confirm vendor approval, banking continuity, contract scope and suspension status.
5	Completion evidence	Check technician report, photographs, part records, sensor recovery and work-order closure.
6	Contract and warranty	Compare rates, premiums, markups, taxes and warranty coverage.
7	Producer recommendation	Assemble the evidence into Pay, Hold or Escalate with uncertainty.
8	Critic attack	Construct the strongest alternative explanation and identify decisive missing evidence.
9	Independent verification	Recalculate material comparisons, verify sources and confirm policy compliance.
10	Accountable decision pack	Present the human controller with the recommendation, conflicts, required evidence and permitted next actions.

Suppose the portfolio finds that the authorized purchase order was C$12,400, a related C$6,300 invoice may already have been submitted, the banking destination changed recently, completion evidence is incomplete and one component may remain under warranty. The evidence-bearing recommendation is **HOLD AND ESCALATE**. The candidate still cannot move money.

23.3 Authority

An admitted Invoice Integrity Successor Ω might receive A3 reversible bounded authority only for temporary holds below C$25,000, with expiry, audit, human override and rollback. It remains prohibited from transmitting payment, changing bank details, creating vendors, modifying contracts, exceeding the ceiling or expanding its own permissions.

24 Worked Example II : Building Operations and the Algorithm Portfolio

24.1 Mission

A portfolio owner seeks a system that can understand the evolving condition of buildings, predict failures, diagnose causes, select interventions, coordinate bounded work and learn across the portfolio without assuming safety, legal or major-capital authority.

The environment is attractive for model-based planning because the true equipment state is partially observed, diagnostic actions change future information, interventions create delayed consequences and digital twins can support repeated episodes. It is also dangerous because a simulator may omit the physical or human variable that matters most.

24.2 Candidate architectures

The Foundry should preserve at least :

— repaired deterministic maintenance rules ;
— behaviour cloning from accepted technician trajectories ;
— model-free sequential RL ;
— MuZero or Sampled MuZero with a learned latent world model and search ;
— Dreamer-style imagined rollouts for continuous or visual control ;
— a sovereign hybrid using deterministic safety rules, specialist intelligence, human engineers and independent verification.

24.3 One episode

The hidden condition is an intermittent control-board failure. The candidate observes tenant complaints, temperature anomalies, uncertain pressure readings, cold weather and a prior valve repair. It may inspect controller logs, request an independent measurement, dispatch a technician, replace a component, monitor, abstain or escalate.

A useful policy discovers that controller-log inspection plus one independent measurement resolves the causal ambiguity before an expensive intervention. The reward vector credits evidence, correct diagnosis, lower cost and reduced downtime. A safety-system override remains a blocked hard failure rather than a tradeable negative reward.

24.4 Proof

One exact release is frozen and evaluated against the incumbent, rules, external property software and a strong human-AI alternative on unseen building cases. Promotion requires zero critical safety misses, zero unauthorized actions, sufficient evidence, improved first-time repair, lower complete burden, transfer across properties and successful rollback. A synthetic formation win is not enough.

Worked-example distinction

The Gym trains candidate policies and architectures. Fresh protected reality determines whether the complete institution has earned Specialist ASI. Accountable principals decide whether it may become Successor Ω.

PART V

The Successor to the Successor

Recursive improvement without recursive self-authorization

The verified recursive operator, Chronicle memory and the institution that manufactures the next admitted release.

SUCCESSOR MANIFOLD → GRADIENT → SEIZE → BOUNDED JOBS → SPECIALIST ASI → SUCCESSOR Ω

25 Recursive Institutional Improvement and Successor Recursion

25.1 Recurrence is not recursive improvement

A system that repeats an earlier workflow is recurrent. A system that modifies its own prompt is self-modifying. Neither condition establishes recursive institutional improvement. The stronger GoalOS definition is :

Definition 25.1 (Verified recursive institutional improvement). An institution recursively improves when a Chronicle-admitted capability measurably increases its ability to detect, constitute, prove or govern a superior successor on fresh work, without reducing proof quality, rights, resilience or governance integrity.

Let generation G_t contain Mission Gym $\mathscr{G}_t$, Successor S_t, Chronicle C_t, proof infrastructure P_t and authority state A_t. The recursive successor operator is

$$(G_{t+1}, \mathscr{G}_{t+1}, S_{t+1}, C_{t+1}, P_{t+1}, A_{t+1}) = \mathfrak{R}(G_t, \mathscr{G}_t, S_t, C_t, P_t, A_t; E_t). \tag{169}$$

The operator is not executed by the candidate alone. It is an institutional process containing formation, independent proof and accountable admission.

25.2 The central recursion

The cycle is

$$S_t \xrightarrow{\text{Chronicle + Gym episodes}} \{S_{t+1}^{(1)}, \dots, S_{t+1}^{(k)}\}, \tag{170}$$

$$\{S_{t+1}^{(i)}\} \xrightarrow{\text{SEIZE + Foundry}} \widehat{S}_{t+1}, \tag{171}$$

$$\widehat{S}_{t+1} \xrightarrow{\text{Fresh-Proof Gate}} S_{t+1}^{\star}, \tag{172}$$

$$S_{t+1}^{\star} \xrightarrow{\text{accountable admission}} S_{t+1}^{\Omega}. \tag{173}$$

The current successor may help generate descendants, identify failure clusters, propose new tools or improve the Gym. It may not certify the descendant, alter protected proof, inherit authority automatically or install the new release.

25.3 Recursion conditions

Promotion from generation t to $t+1$ requires :

$$\mathrm{LCB}\,(V_m(S_{t+1}) - V_m(S_t)) > \theta_{\text{gen}}, \tag{174}$$

$$\Delta\text{ProofQuality} \geqslant 0, \tag{175}$$

$$\Delta\text{GovernanceIntegrity} \geqslant 0, \tag{176}$$

$$\Delta\text{RightsClarity} \geqslant 0, \tag{177}$$

$$\text{UnauthorizedActions} = 0, \tag{178}$$

$$A_{t+1} \preceq \phi(E_{t+1}). \tag{179}$$

A newer architecture is not better because it is larger, faster or more autonomous. It is better only when it produces greater verified mission value at lower justified burden while preserving mission purpose, evidence, reversibility and human sovereignty.

25.4 Recursive gain

Let m_{t+1} be a fresh mission or materially changed environment. Define generation gain

$$\Delta_G(G_{t+1}, G_t \mid m_{t+1}) = a_Q \Delta Q + a_V \Delta V - a_C \Delta C - a_R \Delta R - a_H \Delta H + a_T \Delta T + a_K \Delta K. \tag{180}$$

Promotion requires a positive preregistered threshold. Merely using prior output does not count. The gain must be attributable to reusable environments, verifiers, failure maps, rights templates, tools, integrations or Chronicle knowledge.

25.5 Successor learning curves

The Foundry becomes a platform rather than bespoke consulting when later missions become cheaper, faster or safer after normalizing for difficulty. Define

$$\text{FormationGain}_n = \frac{B_{\text{institution}}(S_1) - B_{\text{institution}}(S_n)}{B_{\text{institution}}(S_1)}. \tag{181}$$

A positive value is valid only if proof and mission difficulty remain comparable. Reduced burden caused by weaker proof is not learning.

25.6 Recursive improvement of the Gym

The Gym itself can improve through :

— better state representations ;
— more realistic transitions ;
— stronger adversarial scenarios ;
— lower-cost replay ;
— improved hidden ground truth ;
— better evaluator calibration ;
— reduced Gym basis risk ;
— transfer cases discovered from production ;
— clearer action and authority schemas.

But the Gym can also be corrupted by the current successor. A candidate may overfit the simulator, shape data collection or suppress failures. Therefore, Gym changes receive their own version, provenance and validation. Let $\mathscr{G}_{t+1} = g(\mathscr{G}_t, E_t)$ be a proposed environment update. It is admitted only if

$$\widehat{\epsilon}_{\text{gym},t+1} < \widehat{\epsilon}_{\text{gym},t} \tag{182}$$

or it expands relevant coverage without weakening protected proof.

25.7 Negative Capability Graph

Chronicle preserves not only winners but also rejected architectures, reward exploits, rights failures, verifier weaknesses, unsafe authority paths and non-transferable improvements. Represent the failure memory as a graph

$$\mathscr{N}_t = (V_F, E_F), \tag{183}$$

where nodes are failure conditions and edges represent shared causes, components, missions or repairs. The graph prevents the institution from repeatedly paying to rediscover the same failure and preserves stepping stones that may become useful under a different regime.

25.8 The recursion firewall

Recursive improvement may expand capability, not self-authorized power. Every descendant starts with no inherited production authority for materially changed claims. Protected cases, scorer internals, release credentials, mission constitution and admission remain outside the candidate's control.

25.9 A bounded monotonicity proposition

Proposition 25.1 (Governance-preserving recursive improvement). *Assume each generation is evaluated on fresh representative work; proof and authority are independently controlled; every material change removes inherited authority for affected claims; and promotion requires positive lower-bound mission gain with no governance regression. Then the admitted sequence can be monotone in verified mission value without being monotone in autonomy.*

Démonstration. By assumption, generation $t+1$ is admitted only when its lower-bound mission value exceeds generation t by the frozen threshold and all hard governance gates are non-decreasing. Authority is separately bounded by current evidence, so an increase in capability does not require an increase in autonomy. Therefore, the sequence of admitted mission-value lower bounds is monotone while the authority sequence may remain constant, contract or expand independently. The proposition does not guarantee such a candidate exists; it constrains admission when one is found. □

25.10 Successor half-life and recursion cadence

Let H_S be the expected half-life of the admitted successor and T_R the time required to generate and prove a replacement. Define

$$\text{RecursiveReadiness} = \frac{H_S}{T_R}. \tag{184}$$

Values below one imply the institution may not generate a credible next successor before the current one becomes impaired. The strategic purpose of reusable Gyms, verifiers and Chronicle is to increase this ratio.

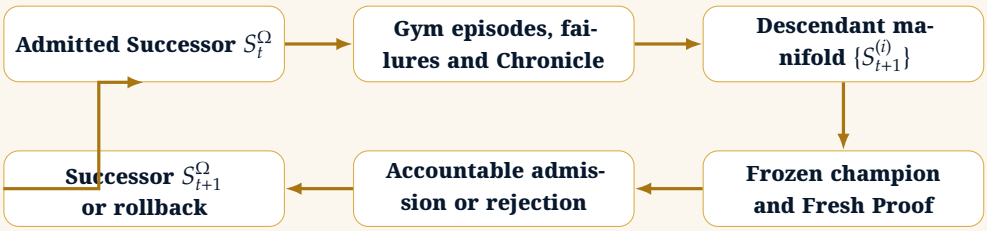


Figure 15. The successor-to-successor recursion. The loop contains independent proof and accountable admission; the candidate cannot close the loop itself.

Recursive succession law

The successor to the successor is not generated by novelty alone. Chronicle proposes; the Foundry diversifies; the Gym exposes; fresh proof selects; authority admits; reality requalifies.

26 The Gym as a Sovereign Corporate Asset

26.1 Environment ownership changes the strategic game

A corporation that rents models but owns no mission environment depends on provider claims, transient prompts and non-portable demonstrations. A corporation that owns a rights-cleared Mission Gym can place providers, models and

architectures into competition against an institution-owned representation of the mission.

The Gym asset includes :

— Mission Constitution and objective history ;
— state, observation and action schemas ;
— rights-cleared historical episodes and synthetic scenario logic ;
— protected fresh cases and case commitments ;
— transition models and digital twins ;
— reward vectors and hard constraints ;
— verifiers, thresholds and failure taxonomies ;
— role and authority simulations ;
— proof receipts and Chronicle interfaces ;
— transfer, canary and requalification environments.

This asset can survive the replacement of a model or vendor. It therefore changes the build–rent boundary.

26.2 The sovereign Gym ledger

Every Mission Gym should carry a ledger :

Table 16. Sovereign Gym Ledger.

Field	Required content
Identity	Gym ID, mission, owner, custodian, version and effective date.
Rights	Data, model, simulator, scenario, output, verifier, reuse and confidentiality rights.
Scope	Valid mission family, populations, regimes, assets, jurisdictions and excluded cases.
State model	Hidden state, observation channels, missingness, uncertainty and transition assumptions.
Actions	Typed actions, masks, permissions, approvals and external-effect classes.
Rewards and constraints	Vector rewards, hard gates, critical failures and non-compensable constraints.
Evidence	Formation, adversarial, transfer, protected and canary partitions ; hashes and custody.
Basis risk	Known reality gaps, external-validity limits, tail omissions and monitoring.
Operations	Runtime, providers, compute, logging, incident response and degraded modes.
Lifecycle	Expiry, revalidation, impairment, rollback, retirement and successor history.
Economics	Cost basis, carrying cost, substitution value, proof value and transfer value.

26.3 Gym sovereignty premium

Let V_G be the value of the Gym asset. A sovereignty premium arises from :

$$\Pi_G = \mathbb{E}[\text{avoided hold-up} + \text{substitution option} + \text{faster proof} + \text{retained learning} + \text{transfer} + \text{authority precision}] - C_G - R_G. \tag{185}$$

The premium is positive only when the environment remains representative and rights-cleared. A poor Gym can lock the institution into the wrong reality as effectively as a poor provider can lock it into the wrong model.

26.4 Build, configure or rent

26.5 Gym portability

A portable Gym uses typed APIs, machine-readable schemas and model-independent proof receipts. Portability requires that a new candidate can be connected without rewriting the mission. Let $\mathscr{I}_{\text{env}}$ be the environment interface and

TABLE 17. The build–rent boundary for mission environments.

Mode	Use when	Required control
Rent Beta	The task is low-differentiation, reversible, low-consequence and adequately solved by the market.	Basic vendor, privacy, quality and exit controls.
Configure workflow	The mission needs context and tools but limited proprietary learning or authority.	Own prompts, mappings, tests and exportable records.
Constitute a successor	The mission is valuable, repeated, differentiable, measurable, proofable and capable of accumulating proprietary evidence or authority.	Own the Mission Constitution, Gym, jobs, proof, Chronicle, permissions, portability and economics.
Preserve reserve	The opportunity may become valuable after a capability, legal, price or data catalyst.	Maintain option triggers and low-cost environment assets without operational authority.

$\mathscr{I}_{\text{candidate}}$ the candidate adapter. The desired separation is

$$\text{Mission logic} \perp \text{provider-specific implementation} \tag{186}$$

up to explicit adapters and known limitations.

Gymnasium provides a standard single-agent environment interface, PettingZoo supports multi-agent environment cycles, and OpenSpiel supports a broad range of game structures (Towers et al., 2024; Terry et al., 2021; Lanctot et al., 2019). GoalOS extends those interfaces with institutional identities, proof custody, authority, Chronicle and succession state.

26.6 Corporate balance-sheet treatment

The Gym should be treated as a strategic intangible asset, not merely test code. Internal accounting should track :

— formation cost;
— maintenance and requalification cost;
— missions and releases supported;
— avoided duplicated proof work;
— provider substitutions enabled;
— failures prevented or detected;
— external or internal licensing value;
— rights and expiry;
— residual basis risk.

This is a managerial valuation, not a claim about financial-statement recognition under a specific accounting standard. The purpose is to prevent the institution from losing the environment, rights and proof history when a project ends or a vendor changes.

26.7 The Gym and corporate bargaining power

A provider knows that an institution without a portable Gym faces high switching cost and uncertain revalidation. An institution with a sovereign Gym can invite providers to prove themselves under the same mission conditions. This creates a contestable successor market.

Let q_v be provider quality, p_v price and s_v switching burden. The institution selects

$$v^\star = \arg\max_v \Big[V_m(v; \mathscr{G}_m) - p_v - s_v - R_v \Big]. \tag{187}$$

Provider power falls as the Gym makes s_v observable and reducible.

Design Proposition 26.1 (Gym ownership). For repeated differentiable missions, the right to define, preserve and re-run the mission environment can be more strategically durable than ownership of any one model. The Gym is the asset that lets future intelligence compete for the mission.

26.8 The Gym must itself be governed

The institution must protect against :

— evaluation leakage ;
— manipulated ground truth ;
— simulator drift ;
— scorer capture ;
— data-rights expiry ;
— selective scenario omission ;
— candidate influence over environment updates ;
— false confidence from synthetic coverage.

Therefore, the Gym has its own custodian, release process, change log, protected partitions, red team, requalification and retirement state.

Gym sovereignty law

The corporation may rent intelligence while owning the world in which that intelligence must prove itself. But the Gym becomes sovereign only when its rights, validity, evidence custody, portability and basis risk are governed as rigorously as the successor it evaluates.

27 Risk Architecture, Proof Debt and the Reality Gap

27.1 Gym basis risk

A candidate can dominate the Gym and fail the real mission. Define

$$\epsilon_{\text{gym}} = \sup_{S \in \mathscr{C}} \left| V_m^{\text{real}}(S) - V_m^{\text{gym}}(S) \right|. \tag{188}$$

The term is not directly observable before deployment. It must be estimated from held-out transfer, live shadow operation, canaries, post-outcome comparison and known omissions.

Sources include :

— incomplete state representation ;
— inaccurate transition dynamics ;
— sparse tail events ;
— unrealistic simulated humans or adversaries ;
— missing legal, social or physical constraints ;
— delayed and confounded outcomes ;
— environment and policy co-adaptation ;
— distribution shift and strategic response.

A conservative promotion statistic is

$$\text{LCB}\left(\Delta_{\text{gym}} - \widehat{\epsilon}_{\text{gym}} - D_P\right) > \theta_m. \tag{189}$$

27.2 Reward hacking and specification gaming

Reward grounding requires that :

— task generators may propose tasks and rewards but may not authorize them ;
— the answer or value must be independently knowable or auditable ;
— the verifier is tested for false acceptance and gaming ;
— relevance, rights and blind-set separation are confirmed ;
— accepted experience retains provenance and scope ;
— reward updates do not rewrite past promotion criteria.

Open-ended search can produce useful stepping stones, but it can also generate tasks that are trivial, impossible or strategically irrelevant. The institution must distinguish open-ended proposal from open-ended authority.

27.3 Proof Debt

Proof Debt is unresolved assumption or missing evidence supporting current valuation, use or authority. Let claim k have consequence weight w_k, uncertainty u_k and current reliance r_k. Define

$$D_P = \sum_k w_k u_k r_k. \tag{190}$$

Proof Debt can grow even while observed performance remains strong if the mission changes, a provider drifts, rights expire or the Gym becomes less representative.

Proof Coverage is

$$\text{ProofCoverage} = \frac{\text{independently supported capability and controls}}{\text{authority, valuation and reuse granted}}. \tag{191}$$

A ratio below one requires reduced authority, additional evidence or both.

27.4 Authority-at-Risk

Define Authority-at-Risk as consequential authority exposed to unproven or expired capability :

$$\text{AaR} = \sum_j w_j A_j \mathbb{1}[\text{ProofCoverage}_j < 1 \ \vee \ \text{Expired}_j = 1]. \tag{192}$$

The objective is not zero autonomy ; it is zero unjustified authority.

27.5 Threat model

The Successor Complex faces threats from :

— prompt injection and malicious data ;
— tool or credential escalation ;
— provider compromise or silent model change ;
— data poisoning and synthetic-label corruption ;
— verifier capture or collusion ;
— evaluation leakage ;
— hidden external calls ;
— unauthorized action paths ;
— corrupted Chronicle admission ;
— supply-chain and infrastructure failure ;
— strategic counterparties exploiting known policy patterns.

The architecture should align with lifecycle risk-management principles such as the NIST AI RMF and Generative AI Profile while adding succession-specific proof, admission and authority objects (National Institute of Standards and Technology, 2023, 2024).

27.6 Twelve canonical failure modes

Failure	What happens	GoalOS control
Reward hacking	Candidate exploits the score rather than the mission.	Independent grounding, hard constraints, adversarial Gym and post-outcome checks.
Environment overfit	Candidate memorizes formation cases.	Hidden Fresh-Proof Gym, transfer families and exact release freeze.
Evaluation leakage	Protected cases enter training.	Independent custody, commitments, access logs and role separation.
Simulator gap	Transition dynamics differ from reality.	Shadow operation, canary, basis-risk reserve and monitoring.
Correlated verifier failure	Producer and verifier share the same blind spot.	Method, model, data and incentive diversity; human adjudication.
Hidden human rescue	Apparent autonomy relies on unreported intervention.	Complete intervention logging and human-burden accounting.
Authority laundering	A high score becomes permission to act.	Separate admission and Authority Lattice.
Action-space escape	Candidate reaches unapproved tools or systems.	External least-privilege gateway and typed tool calls.
False completion	Deadline expiry is recorded as success.	Termination/truncation separation.
Distribution shift	Production leaves validated conditions.	Drift detection, fail-closed modes and requalification.
Environment capture	Candidate or provider influences its own test.	Independent Gym ownership, change control and audit.
False Chronicle admission	Unsafe or stale capability shapes future work.	Admission criteria, expiry, challengeability and revocation.

27.7 Human oversight and professional authority

Human oversight is only meaningful when the reviewer has :

— adequate time and information ;
— authority to refuse ;
— domain competence or access to it ;
— a clear evidence record ;
— independence from the candidate's promotion incentives ;
— a practicable rollback route.

A human click inside a high-pressure workflow is not necessarily accountable admission. The system must measure reviewer burden, disagreement, overrides and rescue.

27.8 Control despite stronger components

The institution should remain safe even if one component is treated as untrusted. Control protocols, restricted tools, audit, monitoring and independent checks can provide useful bounds even when capability exceeds the verifier on some dimensions (Greenblatt et al., 2024 ; Shavit et al., 2023 ; Shevlane et al., 2023 ; Phuong et al., 2024).

27.9 Tail loss

Tail loss includes :

— unauthorized financial or legal action ;
— security compromise or data exposure ;
— harmful physical action ;
— discriminatory or unlawful decision ;
— systemic correlated error ;
— irreversible dependency or institutional capture ;
— false admission into Chronicle ;
— capital allocation based on fabricated value.

Risk law

A correct secular thesis can still produce institutional ruin through bad sizing, weak proof, concentrated dependency, path dependence, Gym basis risk or uncontrolled authority. GoalOS governs the path, not merely the direction.

28 Technical Architecture : From Mission Gym to Successor Operations

28.1 The six-plane architecture

A production implementation separates public orientation, confidential formation, protected proof and institutional authority. The planes are ordered by consequence rather than by visual prominence.

1. **Public Navigator.** Public-safe research, capability radar, education, mission qualification and non-confidential board orientation.
2. **Protected Mission Workspace.** Confidential Mission Constitution, incumbent ledger, economics, rights, proof capital and Authority ceiling.
3. **Sovereign Mission Gym.** Versioned environments, replay, digital twins, action schemas, constraints, scenario generation and proof receipts.
4. **Isolated Successor Foundry.** Data formation, models, agents, tools, search, simulations, candidate formation and sandbox execution.
5. **Independent Proof Plane.** Protected cases, scorer internals, validator credentials, Fresh-Successor Gate, canary evidence and Release Certificate.
6. **Chronicle, Authority and Capital Plane.** Admission, Succession Ledger, monitoring, realized value, impairment, requalification, Capital-to-Capacity and board reporting.

28.2 Environment API

A Gymnasium-compatible core interface is :

```
observation, info = env.reset(seed=42)
while True:
   action = successor.act(observation, authority=info["authority"])
   observation, reward, terminated, truncated, info = env.step(action)
   receipt_store.append(info["proof_receipt"])
   if terminated or truncated:
      break
```

GoalOS extends the environment API with player identity, Authority Envelope, action masks, vector rewards, constraint events, proof receipts, evidence lineage, exact Gym and candidate versions, declared human interventions, Chronicle eligibility, rollback and requalification triggers. Multi-player missions can use PettingZoo-style agent cycles or OpenSpiel-style game representations when roles act sequentially, simultaneously or under imperfect information (Terry et al., 2021; Lanctot et al., 2019).

28.3 Identity, least privilege and evidence custody

Protected identity controls

- organizational identity and role-based access;
- separate development, evaluation, release and audit roles;
- short-lived credentials and scoped sessions;
- encrypted storage and key management;
- data classification and jurisdictional controls;
- append-only audit evidence;
- dual control for consequential release;
- incident containment and credential revocation.

Independent proof custody

- blind-case commitments and hashes;
- scorer versions and frozen thresholds;
- evaluation runtime and environment;
- exact candidate manifest;
- result receipts and confidence calculations;
- reviewer and acceptance identities;
- release signature and rollback package;
- technical separation from candidate formation.

28.4 Machine-readable canonical records

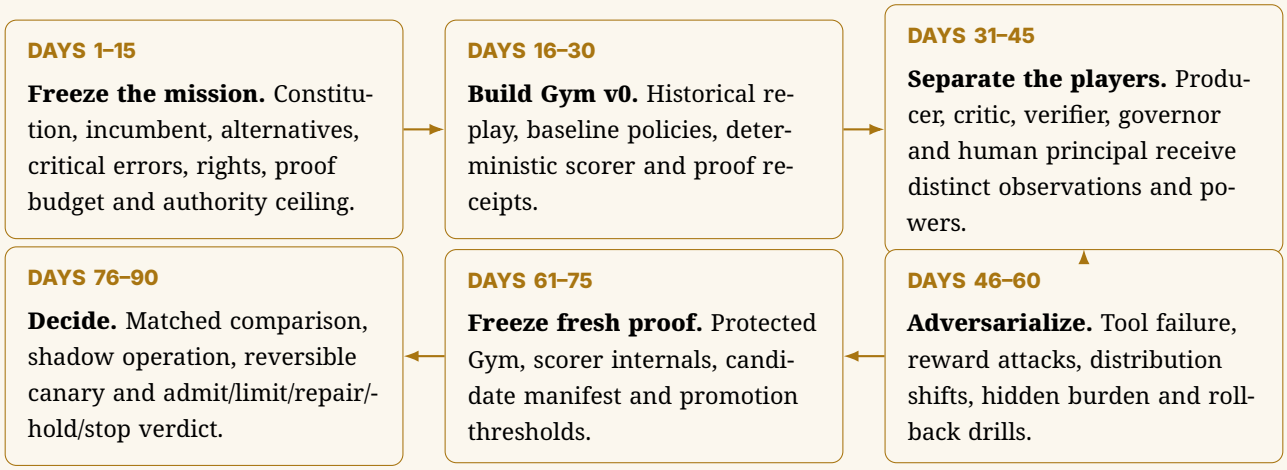
Every state transition carries an attributable actor, evidence object, authority receipt, timestamp, version and stop condition. The minimum record set is :

1. FrontierEventRecord;
2. SuccessionEventRecord;
3. IncumbentAssumptionLedger;
4. MissionGymLedger;
5. EnvironmentSpec;
6. SuccessorBook;
7. SuccessorUnderwriting;
8. ProofCapitalAuthorization;
9. SuccessionConstitution;
10. BoundedAGIJobPortfolio;
11. VerifierSpec;
12. CandidateSuccessorManifest;
13. RunProof;
14. FreshSuccessorScorecard;
15. PromotionRequest;
16. ReleaseCertificate;
17. AuthorityEnvelope;
18. ChronicleSuccessorRecord;
19. CapitalAllocationMemo;
20. RequalificationRecord;
21. ImpairmentRecord;
22. RevocationRecord.

28.5 Operational observability

A production dashboard should expose the current mission and release; Gym and scorer versions; evidence freshness; active Authority Envelope; blocked and attempted unauthorized actions; human interventions and reviewer burden; cost, latency and resource consumption; drift and out-of-distribution alerts; rollback readiness and current fallback; and the next requalification date.

28.6 90-day activation path



28.7 Production hardening

The public browser demonstrations accompanying GoalOS are reference implementations rather than protected production systems. Real deployment requires server-side identity and authorization; confidential evidence storage; independent proof custody; model and provider monitoring; secrets management; protected audit and release signing; security, privacy, legal and professional review; operational continuity and tested rollback; and independent calculation of realized value.

Architecture law

Public navigation may explain succession. Protected workspaces may constitute it. Independent proof may validate it. Only the Chronicle, Authority and Capital Plane may admit, govern and compound it.

29 The Successor Compiler and AGI Job Runtime

29.1 From constitutional text to executable control

The paper becomes operational only when the Mission Constitution, Job Graph, environment, verifier and Authority Envelope can be represented as versioned machine-readable objects. The compiler translates institutional intent into typed executable constraints; the runtime enforces them; the independent proof service determines what passed; the admission service remains human-controlled.

29.2 Compiler pipeline

The canonical compilation path is :

Mission Constitution → Type Check → Job Graph → Capability Masks → EnvironmentSpec → VerifierSpec → Frozen Candidate Manifest.

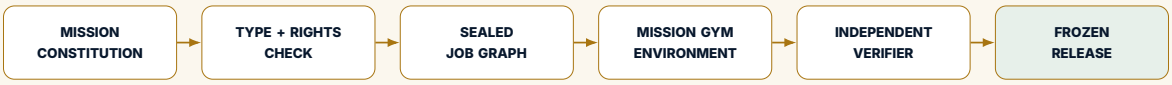


Figure 16. The constitutional compilation pipeline. No stage may silently infer rights, authority or acceptance from upstream capability.

29.3 Runtime services

A production runtime should expose separate services :

Service	Responsibility
Constitution Registry	Stores versioned missions, principals, prohibitions, hard gates, evidence standards and expiry.
Job Scheduler	Releases only jobs whose typed dependencies, rights, budget and authority masks are currently valid.
Capability Gateway	Enforces least privilege over models, tools, data stores, accounts, networks and external actions.
Evidence Service	Preserves source lineage, receipts, hashes, human interventions, costs and accepted outcomes.
Verifier Plane	Holds protected cases, scorer internals and release credentials outside candidate control.
Admission Service	Records accountable human acceptance, Authority Envelope, expiry, monitoring and rollback.
Chronicle Service	Admits only scoped, rights-cleared, versioned and challengeable capability and failure records.
Requalification Service	Invalidates affected claims after material change and triggers new proof or authority contraction.

29.4 Capability-secure execution

Each job identity receives an unforgeable capability set rather than ambient access. The runtime checks

$$\mathrm{Allow}(i, a, o, t) = \mathrm{Cap}(i, a, o) \wedge \mathrm{Current}(t) \wedge \mathrm{WithinBudget}(i), \tag{193}$$

where i is identity, a action, o object and t current evidence and expiry state. Denied actions are recorded as proof-relevant events. A model instruction cannot override the capability gateway.

29.5 Event sourcing and proof receipts

The runtime is event-sourced. Every consequential transition appends an immutable record containing :

— mission, job, candidate and environment versions ;
— acting identity, principal and capability mask ;
— authorized inputs, tools and external dependencies ;
— action, output, retries, human rescue and complete cost ;
— verifier status, residual uncertainty and blocked actions ;
— rollback target, expiry and cryptographic digest.

The current institutional state is reconstructed from the accepted event stream, not from mutable narrative summaries.

29.6 Execution algorithm

Algorithm 3 Execute one sealed AGI Job

Require : signed job J, current mission state M_t, capability gateway C, evidence service E, verifier V
Ensure : accepted receipt, typed refusal or rollback

```
verify signature, version, rights, dependencies, budget and expiry
if any precondition fails then
    emit typed refusal and stop
end if
issue short-lived job identity and least-privilege capability mask
execute bounded work while recording every action and external dependency
if prohibited action attempted or hard constraint violated then
    revoke capabilities; execute rollback; emit failure receipt; stop
end if
seal output, cost, interventions and evidence lineage
submit output and receipt to independent verifier V
if acceptance predicate fails then
    route to repair, reject or rollback branch; stop
end if
append accepted proof receipt to Evidence Service
mark output Chronicle-eligible only if rights, scope and expiry allow
return accepted receipt
```

29.7 Fresh-proof release protocol

The proof plane receives only the signed frozen manifest and a one-way execution interface. It does not expose protected cases, score explanations or gradients to the candidate. The protocol is :

CASE COMMITMENT → ONE-WAY EXECUTION → INDEPENDENT SCORING → SIGNED VERDICT → ACCOUNTABLE REVIEW.

The result may be promoted, retained, repaired, reserved, hedged, rejected or stopped. Only the accountable review may issue a Release Certificate or Authority Envelope.

29.8 Operational observability

The board and operators should be able to observe, by exact release :

— current mission and Authority Envelope ;
— jobs running, blocked, failed, repaired and awaiting verification ;
— proof coverage by claim and action class ;
— critical-error, abstention and unauthorized-action counts ;
— human interventions, cost, latency and provider dependence ;
— current rollback target and degraded operating mode ;
— evidence expiry and next requalification date.

29.9 Production boundary

A public GitHub Pages interface can explain the institution, compile non-confidential local objects and export Mission Packs. It cannot safely hold confidential evidence, protected scorers, production credentials or GPU-scale training. Consequential deployment requires isolated formation, independent proof custody, server-side identity, encrypted evidence, release signing, incident response and tested rollback.

Runtime law

The compiler translates the constitution into capabilities and evidence obligations. The runtime may execute sealed work. The proof plane may sign a verdict. Only accountable authority may create a Successor Ω release.

30 Paper-to-Code Reproducibility and the Evidence-Bearing Publication

A paper about verified succession should itself expose a reproducible release architecture. Version 3.0 therefore treats publication as a bounded evidence object : exact source, deterministic synthetic benchmark, generated figures, environment cards, claim boundary, build logs, manifest and cryptographic checksums are distributed together.

30.1 Reproducibility planes

Table 20. Publication reproducibility planes.

Plane	Authoritative object	Reproduction test
Scholarly source	Modular and single-document LaTeX, bibliography and exact figures.	Clean-room A4 and US Letter compilation with resolved references and embedded fonts.
Synthetic evidence	GSSB-4 generator, fixed seed, case-level results, environment cards and gate thresholds.	Regenerate 6,000 cases, 30,000 episodes, tables, figures and hashes.
Institutional visuals	Deterministic vector plates plus supplementary image-generated ornament.	Verify that operative text and equations exist independently of image microtypography.
Machine-readable constitution	Environment, candidate, proof, authority, Chronicle and requalification schemas.	Validate records against schemas and replay the canonical state transition.
Distribution	PDFs, HTML, EPUB, source archives, manifest and SHA-256 ledger.	Extract every archive, compare payload hashes and audit internal links.

30.2 Paper-to-code crosswalk

Paper object	Executable object	Evidence emitted
Mission Gym $\mathscr{G}_m$	Environment card and case generator	Versioned state, observations, actions, constraints, split identity and case commitments.
Successor Manifold $\mathscr{M}_S$	Candidate profile and architecture manifest	Models, tools, verifier, governance, sovereignty, cost and authority coordinates.
Mission Advantage Gradient	Candidate mutation and bounded search signal	Proposed architecture delta, expected value direction and proof-capital request.
SEIZE Underwriting	Gate threshold and option decision record	Mission, alternatives, proof budget, stop-loss, challenger freeze and review owner.
Bounded AGI Jobs	Typed work-and-proof record	Inputs, tools, prohibitions, budget, verifier, acceptance and rollback.
Fresh-Successor Gate	Paired proof evaluator	Case-level utility differences, confidence bounds, critical errors and gate verdict.
Authority Envelope	Machine-readable permission state	Scope, ceiling, expiry, blocked actions, rollback and accountable principal.
Chronicle	Append-only successor record	Accepted capabilities, failures, proof state, release history and requalification trigger.

Paper object	Executable object	Evidence emitted
Recursive operator $\mathfrak{R}$	Generation table and lineage archive	Descendants, rejected lineages, basis-risk reserve, champion and unchanged authority ceiling.

30.3 One-command synthetic rebuild

The Research Companion contains the deterministic benchmark runner. From its directory :

```
python run_gssb4.py --out generated
```

The command recreates :

— four mission datasets and three partitions per mission ;
— case-level outputs for five complete candidate architectures ;
— candidate summaries, paired-gain statistics and gate verdicts ;
— recursive-generation records ;
— four publication figures ;
— environment cards and the benchmark manifest.

The publication build then compiles A4, US Letter, Board Brief, Visual Abstract, Evidence Dossier and Institutional Atlas editions. The source package includes a single build entry point, while the release suite includes clean-room reconstruction records so that compilation success is not inferred from the author's working directory.

30.4 Claim ledger

Every material claim in the release belongs to one of five classes :

1. **Definition** : a stipulated GoalOS term or constitutional object ;
2. **Derived result** : a theorem or proposition following from explicit assumptions ;
3. **External fact** : a statement attributed to a cited source ;
4. **Synthetic result** : a deterministic benchmark or simulation output ;
5. **Real-world state claim** : a claim requiring exact customer evidence, independent calculation and accountable admission.

The release never promotes a synthetic result into the fifth class. The accessible HTML edition and Evidence Dossier repeat this boundary so that distribution format does not change epistemic status.

30.5 Environment and dataset cards

Each synthetic mission has an environment card specifying :

— mission, beneficiary and reference architecture ;
— synthetic state variables and generation process ;
— observation and action families ;
— formation, proof and transfer partitions ;
— hard constraints and critical failures ;
— candidate visibility and protected elements ;
— known simplifications and Gym basis risks ;
— authority ceiling and claim boundary.

The cards are designed to be replaced, not merely edited, when a rights-cleared customer environment is constituted.

30.6 Accessibility and durable distribution

The ceiling release includes a semantic HTML edition and EPUB companion in addition to print PDFs. Equations are represented in source form, figures include descriptive captions, tables retain header structure and the reading order follows the scholarly hierarchy. The accessible editions are not a substitute for professional accessibility certification, but they avoid making the fixed-layout PDF the only usable source.

30.7 Publication as a proof-carrying object

The release manifest identifies each artifact, its size, SHA-256 digest, role and relation to the authoritative source. The complete suite can therefore be treated as a proof-carrying publication :

$$\mathscr{P}_{\text{release}} = (\text{source}, \text{build}, \text{evidence}, \text{figures}, \text{claims}, \text{manifest}, \text{checksums}, \text{verification}). \tag{194}$$

A reader may disagree with the assumptions, utility function or synthetic world model while still reproducing exactly what the paper did.

Reproducibility law

A succession paper should not ask the reader to trust an invisible pipeline. Definitions must be inspectable, synthetic results reproducible, protected claims bounded, and every distributed artifact attributable to one versioned source of truth.

31 Corporate Productization : Mission-Sovereign Successor-as-a-Service

31.1 The corporate business model

MONTREAL.AI develops and commercializes the institutional system for constituting Mission-Sovereign Successors. It combines rented frontier intelligence with a customer's mission, proprietary knowledge, data, tools, workflows, verifiers, evidence, controls and human authority—then charges to underwrite, constitute, prove, deploy, monitor and continuously renew the resulting institutional capability.

$$\boxed{\begin{array}{l}\text{Rented AI Beta} + \text{mission-owned complements} \\ \qquad + \text{GoalOS Gym, proof and governance}\end{array} \longrightarrow \text{Mission-Sovereign Successor.}} \tag{195}$$

The company does not primarily sell access to a model. It sells the corporate capacity to manufacture, prove, govern and continuously renew the superior successor for a consequential mission.

31.2 One flagship business, two commercial tracks

Low-cost education, workflow experimentation and proof starters. The Lab generates demand, shared language, use cases, buyer evidence and enterprise-qualified missions.

The flagship enterprise progression. It sells capital-allocation decisions, sovereign Mission Gyms, candidate formation, protected proof, bounded admission and recurring successor operation.

31.3 GoalOS Lab : acquisition and qualification layer

Offer	Customer outcome	Illustrative price
Public Navigator	Public research, capability radar and succession qualification.	Free
GoalOS Lab product	Build one reusable or self-improving workflow and learn the proof discipline.	US$49–997
Executive Successor Workshop	Select one consequential mission, map the incumbent and define a preliminary Spread and SEIZE mandate.	US$10,000–25,000
Board Successor Briefing	Executive exposure map, one board-ready opportunity memo and a decision on the next bounded mandate.	From US$15,000

31.4 Enterprise successor ladder

Offer	Institutional deliverable	Illustrative global price
Singularity Readiness Mission	Capability Twin, incumbent ledger, Gymmability assessment, opportunity portfolio and 30/90/365-day Successor Book.	US$50,000–100,000
SEIZE Succession Mandate	Frozen mission, Successor Manifold, Mission Advantage Gradient, complete underwriting, proof-capital plan and go/hold/-repair/stop decision.	US$125,000–300,000
Mission Gym and Formation Pilot	Sovereign Gym v0, competing complete architectures, Bounded AGI Jobs, independent verifier and one exact frozen challenger.	US$350,000–1,000,000
Verified Mission Pack	Accepted architecture, Evidence Docket, Release Certificate, Authority Envelope, rights, rollback, monitoring and complete operating economics.	US$750,000–2,000,000
Mission-Sovereign Successor Operations	Hosted or protected operation, evidence custody, Chronicle, monitoring, incident controls, requalification and provider substitution.	US$250,000–1,000,000 per mission annually plus usage
Enterprise Successor Portfolio	Multiple mission successors sharing identity, Gym infrastructure, policy, Chronicle, proof services and board reporting.	US$1.5M–5M annually
GoalOS Singularity Office Ω	Permanent succession office : radar, Gym portfolio, Foundry, Independent Proof Plane, Chronicle and Capital-to-Capacity.	US$3M–10M+ annually

The figures are strategic price bands, not promises or market evidence. Exact terms depend on mission consequence, data sensitivity, proof difficulty, integrations, Authority Envelope, jurisdiction and service levels.

31.5 Sell the decision before the build

The first paid decision

Is the incumbent materially impaired, which successor architecture is most promising, what is the cheapest decisive proof, what is the maximum downside, and should the institution fund, hold, hedge, partner, repair or stop?

This protects both parties. The customer does not fund a large implementation before the mission, rights and proof path are credible. MONTREAL.AI does not guarantee an outcome that depends on customer data, systems, budgets and authorities that have not yet been frozen.

31.6 Recurring revenue is structurally justified

A successor is not proven forever. Models, providers, data, laws, missions, scorers and authority change. Recurring revenue can therefore be tied to actual renewal evidence :

- frontier changes detected early enough to change the customer decision ;
- incumbent assumptions correctly impaired or preserved ;
- successor opportunities tested, admitted or rejected ;
- unsafe or uneconomic formations avoided ;
- Proof Debt retired ;
- value realized ;
- fresh transfer achieved ;
- productive Gym, verifier and Chronicle capacity created.

A fixed strategy document expires. A succession institution earns renewal by keeping the architecture current.

31.7 The Mission Pack delivery object

The canonical Verified Mission Pack contains :

- Mission Constitution and preservation constraints ;
- Succession Event and incumbent impairment record ;
- Sovereign Gym Ledger and protected proof plan ;
- Successor Manifold and underwriting book ;
- Bounded AGI Job portfolio ;
- exact candidate manifest and Release Certificate ;
- Authority Envelope and blocked actions ;
- Evidence Docket and proof receipts ;
- rights, privacy, security and professional boundaries ;
- monitoring, expiry, requalification and rollback ;
- Succession Ledger ; Capital-to-Capacity memo.

31.8 Corporate architecture

The intended architecture separates institutional functions :

- **MONTREAL.AI** : parent institution, research, standards, strategic authority and global communications ;
- **QUEBEC.AI** : commerce, checkout, digital products, workshops, implementation and buyer delivery ;
- **GoalOS** : the successor operating system, Mission Gym, Foundry, proof and Chronicle architecture ;
- **project-specific ventures** : separately constituted applications and operating entities with exact contractual scopes.

31.9 Why the Gym strengthens the moat

A model provider can copy features. A competitor can imitate prompts. It is harder to reproduce a rights-cleared Gym containing historical episodes, protected cases, failure maps, verifiers, authority patterns, transfer evidence and Chronicle history. The moat therefore shifts from model access to the integrated mission environment and succession record.

Commercial doctrine

GoalOS Lab creates demand. Readiness creates deal flow. SEIZE sells the decision. The Mission Gym makes the mission executable. The Foundry constitutes candidates. Fresh Proof establishes Specialist ASI. Successor-as-a-Service creates recurring revenue. The Singularity Office compounds a portfolio of Mission-Sovereign Successors.

32 Implementation Blueprint : From Navigator to One Admitted Successor

32.1 The ninety-day activation path

The unified institution should be activated through one narrow, valuable and proofable mission rather than through enterprise-wide automation.

Period	Institutional work	Required decision artifact
Days 1–15	Freeze mission, beneficiary, incumbent, best alternatives, critical failures, proof budget, rights, authority ceiling and review date. Build the Capability Twin and Incumbent Ledger.	Mission Constitution, Succession Event Record and first Successor Book.
Days 16–30	Build Replay Gym v0, baseline policies, deterministic or independently knowable scorer, proof receipts and rights-cleared case set.	EnvironmentSpec, VerifierSpec and incumbent baseline.
Days 31–45	Add producer, critic, verifier, governor and human-principal roles with separate observations, permissions and incentives. Generate materially different candidate institutions.	Candidate manifests, anti-recommendations and Bounded AGI Job portfolio.
Days 46–60	Adversarialize the Gym : tool failure, corrupted evidence, distribution shift, reward attacks, provider outage, human absence and rollback drills.	Failure taxonomy, Negative Capability Graph and revised proof plan.
Days 61–75	Freeze one challenger, protected cases, scorer internals, thresholds, cost denominator and promotion rules.	Promotion Request, cryptographic release identity and independent custody receipt.
Days 76–90	Run matched fresh proof, shadow operation and one reversible canary ; then admit, limit, repair, retain or stop.	Fresh-Successor Scorecard, Admission or Rejection Record, Authority Envelope and Chronicle disposition.

32.2 Mission selection gate

The first mission should normally satisfy :

— high value or high frequency ;
— a clear incumbent and strong external alternative ;
— a bounded state and action space ;
— a rapid, trustworthy and independently governable evaluator ;
— rights-cleared examples, failures and feedback ;
— measurable complete cost and human burden ;
— a realistic path to shadow, sandbox or reversible canary operation ;
— enough recurrence for Chronicle to improve the next cycle.

A mission should be held or rejected when no credible evaluator exists, the opportunity expires before proof, data rights are inadequate, the best external alternative is already superior, or the consequence of error exceeds the available control architecture.

32.3 One real succession cycle

A complete cycle produces, in order :

1. an independently supported incumbent baseline and impairment thesis ;
2. a Successor Manifold containing at least three materially different complete architectures ;
3. a Mission Advantage Gradient identifying the highest-value architecture change ;
4. a SEIZE decision with proof-capital ceiling, stop-loss and decisive evidence ;
5. a typed portfolio of Bounded AGI Jobs ;
6. one frozen challenger and one protected matched comparison ;
7. a mission-dominance verdict or evidence-backed rejection ;
8. a separate accountable authority decision ;
9. bounded canary evidence and realized-value measurement ;
10. Chronicle admission, impairment or rejection ;

11. one material change followed by fresh requalification;
12. a next-successor result that does not inherit proof or authority automatically.

The synthetic Aurelia and GSSB-4 programmes demonstrate the mechanics of this loop. A real completion requires rights-cleared customer data, independent proof custody and accountable operational authority.

32.4 Illustrative acquisition-underwriting cycle

For a corporation that repeatedly evaluates acquisitions, the constituted mission is to produce an evidence-bearing recommendation that identifies material value drivers, contradictions, downside scenarios, integration risks and decision conditions within a defined diligence window, while reserving all binding authority to the investment committee.

The initial manifold may include :

1. an improved incumbent process;
2. a general frontier-model copilot;
3. a specialist finance–legal–technology–market cell;
4. a thesis/anti-thesis institution with independent verification;
5. a sovereign hybrid using confidential local processing, deterministic financial models, selected frontier reasoning and an institution-controlled evidence service;
6. an acquired external diligence platform.

The Mission Advantage Gradient may point away from better prose and toward source reconciliation, contradiction discovery, active information acquisition, integration simulation, independent evidence coverage and lower senior-review burden. Initial authority remains Observe, Advise and Prepare. The investment committee retains external communication, contract, condition-waiver, capital and transaction authority.

Illustrative proof tranches are shown in Table 24.

Table 24. Illustrative acquisition-underwriting proof tranches.

No.	Question	Evidence	Decision
0	Can the candidate reconstruct source-grounded facts without leakage?	Historical data rooms with hidden answer keys and exact provenance.	Stop on critical-source failure.
1	Does adversarial architecture detect more material contradictions?	Matched cases against incumbent and strongest copilot.	Continue only if gain exceeds review burden.
2	Do complete economics remain positive?	Timed shadow operations, compute, professional review, exceptions and rework.	Narrow or stop if burden exceeds value.
3	Is bounded live use safe and useful?	One shadow or canary transaction with no binding external action.	Admit only the lowest justified authority level.

32.5 First-year operating target

By year end, a credible institution should operate :

— continuous Succession Radar;
— weekly Opportunity and Risk Brief;
— monthly Successor Book and capital committee;
— quarterly institutional reconstitution;
— material-event Capability Shock Protocol;
— at least one sector-specific Mission Pack;

— independent evaluators and replayable benchmarks;
— realized-value and Succession Ledgers;
— at least one fresh-successor comparison;
— one Capital-to-Capacity allocation;
— zero material unauthorized actions.

32.6 The empirical completion gate

The framework should not be declared empirically complete until it has produced :

1. several paid succession mandates;
2. one independently supported Mission-Dominant Specialist ASI;
3. one separately accountable Mission-Sovereign admission;
4. one evidence-backed rejection that prevented material waste or risk;
5. one independently calculated realized-value result;
6. one repeat or expansion customer;
7. one fresh-transfer gain on a new mission;
8. one requalification, impairment or revocation;
9. evidence that Chronicle improved a later mission;
10. zero material unauthorized actions.

Implementation law

Install the constitution before the autonomy. Sell the decision before the build. Freeze the challenger before the proof. Separate dominance from authority. Measure value after the complete denominator. Requalify every material change.

33 Falsifiability, Empirical Programme and the Completion Gate

33.1 What would falsify the framework

The Gym–Specialist ASI–Successor architecture should be revised or rejected if evidence shows that :

— the Mission Gym does not improve decisions relative to simpler governance ;
— environment formation and maintenance costs systematically exceed avoided loss or realized value ;
— Gym performance fails to transfer after basis-risk adjustment ;
— proof costs are consistently greater than the value of reduced uncertainty ;
— quality-diversity archives add delay without useful option value ;
— the complete Successor rarely outperforms simpler human or software architectures ;
— Chronicle does not reduce repeated work or repeated failure ;
— later missions are not cheaper, faster or safer after normalizing for difficulty ;
— authority separation is routinely bypassed ;
— successors cannot be revoked or unwound safely ;
— customers do not renew based on realized evidence ;
— provider substitutability does not create practical value.

Clean negative findings strengthen the system. A proof-backed decision to retain, rent, narrow or stop is part of the value proposition.

33.2 Preregistered hypotheses

H	Hypothesis	Falsification test
H1	A sovereign Mission Gym reduces time-to-verified-successor.	Compare matched missions with and without reusable environment infrastructure.
H2	Complete-successor evaluation predicts production value better than model-only benchmarks.	Regress realized outcomes on model scores versus complete institutional scorecards.
H3	Independent verifier formation reduces critical false reassurance.	Randomize verifier independence and compare protected failure rates.
H4	Quality diversity improves robustness and transfer.	Compare archive-based Foundry with single-leader optimization on fresh regimes.
H5	Proof-capital staging reduces wasted formation spend.	Compare staged SEIZE decisions with full-build-first programmes.
H6	Gym sovereignty reduces provider switching cost and hold-up.	Measure requalification burden after controlled provider substitution.
H7	Chronicle reduces formation burden on later missions.	Estimate normalized FormationGain across mission generations.
H8	Authority Lattices reduce unauthorized action without destroying mission value.	Compare bounded and collapsed authority designs under matched canaries.
H9	Successor recursion produces positive fresh-transfer gain.	Test successor-to-successor releases on protected new mission families.
H10	Recurring customers renew based on realized succession value.	Link renewal to accepted value, avoided risk, transfer and proof retirement rather than narrative satisfaction.

33.3 Synthetic ceiling and empirical boundary

Aurelia and GSSB-4 satisfy the synthetic reference gate : the environments, candidates, case-level outputs, proof statistics, authority states and recursive generations can be reproduced from disclosed code and fixed seeds. They do not

satisfy the real-world completion gate. In particular, synthetic transition models cannot establish actual rights, safety, organizational adoption, provider behaviour, realized customer economics or production authority.

The real empirical programme should preserve the same state machine while replacing the synthetic environment with one rights-cleared mission. The protected proof custodian, candidate freeze, matched reference set, hard critical-error gates, complete denominator and accountable admission must remain at least as strict as in the synthetic reference.

33.4 Minimum real-world empirical proof

A credible commercial and constitutional completion should include :

1. several paid succession mandates ;
2. one independently supported Mission-Dominant Specialist ASI ;
3. one accountable Mission-Sovereign admission decision ;
4. one evidence-backed rejection that prevented material waste or risk ;
5. one independently calculated realized-value result ;
6. one repeat or expansion customer ;
7. one fresh-successor gain on a new mission ;
8. one requalification, impairment or revocation decision ;
9. evidence that a later mission benefited from Chronicle and shared Gym infrastructure ;
10. zero material unauthorized actions.

33.5 The next constitutional generation gate

The institution should not declare a new constitutional generation until the full loop is demonstrated :

$$\begin{aligned}\text{Succession event} &\to \text{bounded proof mandate} \to \text{frozen challenger} \to \text{fresh proof} \\ &\to \text{human release} \to \text{realized value} \to \text{Chronicle admission} \\ &\to \text{superior next successor.}\end{aligned} \tag{196}$$

33.6 Research agenda

Open questions include :

— How should Gym basis risk be estimated before production outcomes exist ?
— Which mission families support valid synthetic experience without simulator theatre ?
— How can verifiers remain adequate when the candidate becomes stronger than individual reviewers ?
— What quality-diversity descriptors best preserve future successor options ?
— How should credit and capital be allocated among models, tools, data, humans and environment assets ?
— What governance mechanisms resist collusion among producer, verifier and provider ?
— How can protected proof remain fresh when candidate iteration is extremely rapid ?
— Which mission-level invariants define provider-gauge equivalence ?
— When does recursive successor formation produce diminishing returns or instability ?
— How should social and public externalities constrain positive firm-level Mission Alpha ?

33.7 Relation to broader ASI pathways

The Gym–Successor architecture is compatible with multiple routes toward advanced intelligence. Scaling can improve components ; paradigm shifts can expand the manifold ; recursive improvement can accelerate descendant formation ; multi-agent collectives can constitute specialist institutions. The architecture does not assume which pathway dominates. It requires each candidate to enter through the same mission, proof and authority gates.

Empirical completion law

The next legitimate threshold is not another conceptual layer. It is one real, rights-cleared succession cycle that proves a complete candidate on fresh work, grants bounded authority, measures realized value, survives requalification and produces a superior next successor with zero material unauthorized actions.

34 The Twelve Laws of Successor Ω × AGI Jobs

1. **No successor by narrative.** Model scale, benchmark reputation, generated explanation and strategic urgency cannot substitute for exact mission evidence.
2. **No work without a principal.** Every consequential AGI Job has a named accountable owner, acting identity, budget, expiry and escalation route.
3. **No task without a proof obligation.** Every accepted output must carry evidence sufficient for an independent verifier to challenge it.
4. **No hidden action space.** Tools, external calls, network paths, identities, accounts and prohibited actions must be explicit.
5. **No claimant as final judge.** Producer, critic, verifier, governor, acceptor and admitter retain separated powers.
6. **No certification by the teaching Gym.** Formation may adapt; protected proof is sealed, one-way and non-learning.
7. **No Alpha without the complete denominator.** Formation, compute, integration, human rescue, proof, maintenance, dependency, requalification and unwind all count.
8. **No Specialist ASI by training score.** Specialist ASI is earned only through fresh comparative proof against the incumbent and best credible alternative.
9. **No authority by capability.** Authority remains collateralized by fresh evidence and capped by the mission constitution.
10. **No sovereignty without substitutability.** The institution must retain mission, evidence, Chronicle and authority when a model or provider changes.
11. **No inheritance of proof or authority.** Descendants may inherit knowledge, tools and environments; every material release earns proof and authority again.
12. **No durable Alpha without renewal.** Chronicle, impairment, re-underwriting, new AGI Jobs, fresh proof, rollback and revocation are permanent functions.

35 Conclusion : Manufacture the Institution, Not Merely the Model

The beginning of Singularity Time does not merely increase the number of intelligent tools available to corporations. It shortens the life of accepted architectures and increases the cost of treating a model call as an institution. The enterprise that rents increasingly capable systems without owning the mission constitution, job graph, evaluator, evidence, memory and authority structure may become more productive while remaining strategically undifferentiated and operationally dependent.

The Successor Ω × AGI Jobs architecture provides a third path between passive rental and uncontrolled autonomy. It rents Beta where Beta is sufficient; underwrites the Spread where a meaningful mission opportunity exists; compiles the opportunity into sealed work-and-proof contracts; uses the Sovereign Mission Gym and Alpha Foundry to manufacture diverse complete candidates; freezes one exact challenger; purchases fresh reality through independent proof; and grants only the authority the evidence can support.

The central recursion is institutional. AGI Jobs do not merely execute the present workflow. They create the evidence, environments, verifiers, tools and failure maps from which a better successor can later be formed. Successor Ω does not merely operate. It preserves the exact accepted state, monitors impairment, constrains authority and exposes itself to replacement. Chronicle does not merely store. It decides what may influence the next mission.

Final doctrine

AGI Jobs manufacture the successor. The Gym manufactures experience. The Foundry competes complete architectures. Fresh proof may earn Specialist ASI. Accountable admission creates Successor Ω. Chronicle prepares the successor to the successor. Compound only what survives fresh reality.

A Notation and Formal Objects

The notation is grouped by institutional function so that the formal architecture can be read as a system rather than as an undifferentiated symbol list.

Mission Gym and multiplayer game

Symbol	Meaning
m	Constituted mission or preregistered mission family.
$\mathscr{G}_m$	Sovereign Mission Gym : a constrained, partially observable stochastic game.
$\mathscr{N}$	Player set : principal, incumbent, producer, critic, verifier, governor, affected parties, providers and adversaries.
$\mathscr{X}$	Hidden mission state space.
$\mathscr{O}_i$	Observation space available to player i.
$\mathscr{A}_i$	Action space permitted to player i before authority masking.
$\mathscr{T}$	Mission transition kernel.
r	Vector-valued mission reward.
c	Hard constitutional constraint vector.
ρ	Scenario distribution over normal, edge, adversarial and transfer cases.
τ	Termination and truncation conditions.

Successor architecture and mission physics

Symbol	Meaning
$\gg_m$	Complete Successor Institution for mission m.
π_i	Policy used by one role or component inside the complete successor.
$\mathscr{M}_S$	Successor Manifold of materially reachable complete architectures.
$G_m(x)$	Mission-dependent institutional metric tensor.
$V_m(x)$	Proof-adjusted mission value field.
$\nabla^{G_m} V_m$	Riemannian Mission Advantage Gradient.
γ	Underwritten architecture trajectory.
$\mathscr{S}_m[\gamma]$	Succession action or complete trajectory burden.
Φ_m	Mission-superiority order parameter.
Σ_m	Mission-Sovereign phase indicator.

Alpha, proof and authority

Symbol	Meaning
F_t	Generally rentable frontier capability : AI Beta.
Γ_m	Gross Mission Advantage Spread.
α_m	Residual risk-adjusted Mission Alpha.
D_P	Proof Debt : unresolved assumptions or missing evidence.
E_t	Current evidence state.
A_t	Current Authority Envelope.
$\phi(E_t)$	Maximum authority collateralized by current evidence.
$\widehat{\epsilon}_{\text{gym}}$	Estimated Mission-Gym-to-reality basis risk.
$\mathscr{M}_m(x)$	Robust promotion margin after lower confidence bound, Gym basis risk, Proof Debt and mission hurdle.

Chronicle and recursive succession

Symbol	Meaning
K_t	Chronicle-admitted institutional capability at time t.
$\Re$	Recursive Successor Operator.
H_S	Evidence-valid half-life of the admitted successor.
T_R	Expected time to form, prove and admit a replacement.
H_S/T_R	Recursive Readiness ratio.
S_t	Current admitted successor release.
$\widehat{S}_{t+1}$	Exact frozen challenger selected for fresh proof.
$S^{\star}_{t+1}$	Mission-dominant successor proven on fresh work.
S^{Ω}_{t+1}	Separately admitted Mission-Sovereign successor release.

Notation rule

A symbol may summarize a design object, but it never substitutes for the exact mission, release, evidence, authority, rights, expiry and rollback records required by the operating constitution.

B Canonical State Machine

The canonical state machine is grouped into four constitutional phases. The phases may iterate, but their authority boundaries may not be collapsed.

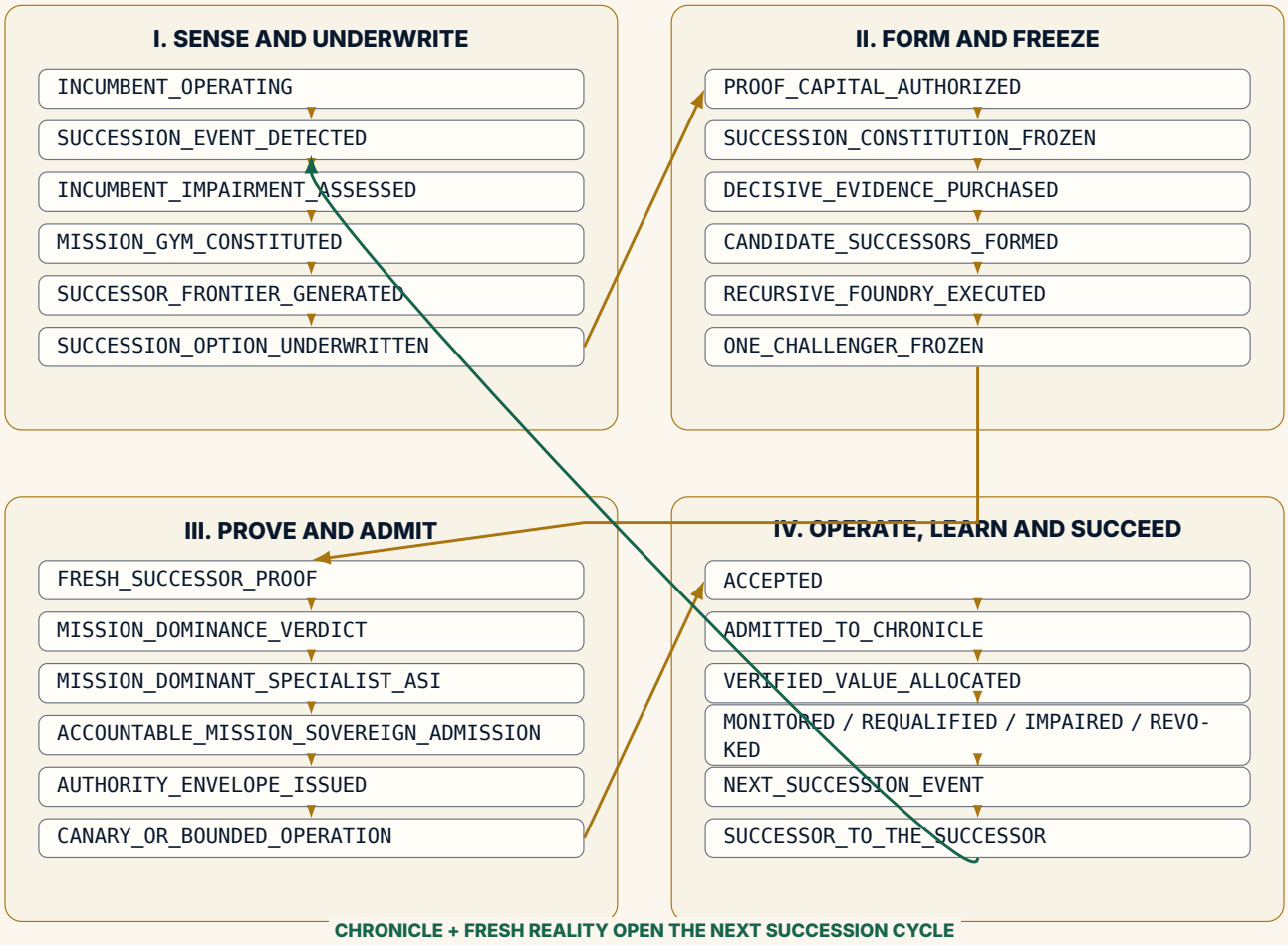


Figure 17. Canonical succession state machine. Every promotion boundary contains evidence, authority and rollback records.

Every transition must include :

— an attributable actor and accountable principal ;
— the exact evidence object and proof receipt ;
— the Mission Gym, candidate and verifier versions ;
— the current Authority Envelope and any blocked action ;
— a timestamp, hash, expiry and stop condition ;
— the known-good rollback target and the next requalification trigger.

State-machine law

No arrow may be inferred from narrative confidence. A state transition exists only when its evidence, authority receipt, accountable actor and rollback route exist together.

C Canonical Twenty-One Bounded AGI Jobs

No.	Job	Required output
1	Mission Constitution	Frozen objective, beneficiary, principal, preserve/build/become constraints, unacceptable failures and prohibited actions.
2	Incumbent Baseline	Accepted quality, cost, risk, time, human burden, evidence and authority state.

No.	Job	Required output
3	Alternative Set	Best credible frontier, specialist, human, hybrid, software, partner, acquisition and no-action baselines.
4	Rights and Data Constitution	Provenance, licences, confidentiality, retention, exclusions, jurisdiction and reuse rights.
5	Mission Gym Constitution	State, observations, actions, transitions, rewards, hard constraints, partitions, verifier and basis-risk plan.
6	Benchmark Constitution	Representative distribution, protected cases, scorer, critical errors and promotion threshold.
7	Oracle and Teacher Study	Declared external intelligence, teaching signal, hidden-call prohibition and transfer hypothesis.
8	Failure Taxonomy	Error classes, severity, causes, abstention requirements, adversarial cases and stop conditions.
9	Architecture Frontier	Diverse complete candidate systems and non-dominated reserve or hedge options.
10	Data and Experience Formation	Rights-cleared examples, simulations, expert demonstrations, synthetic generation and quality controls.
11	Specialist Model Formation	Adaptation, training, distillation or algorithmic construction with exact releases.
12	Tool and Retrieval Formation	Mission-specific tools, deterministic checks, retrieval, planners, sandboxes and external interfaces.
13	Verifier Formation	Independent evaluation harness, calibration, replay, attacks, hidden ground truth and protected scoring.
14	Security and Privacy Proof	Threat model, data boundary, credential controls, isolation, privacy, incident response and provider risk.
15	Complete Economics	Formation, operation, compute, human review, maintenance, requalification, dependency and unwind.
16	Human Burden and Agency	Escalation, contestability, supervision, professional authority, role redesign and reviewer capacity.
17	Resilience and Rollback	Failover, graceful degradation, portability, kill switch, incumbent fallback and known-good restoration.
18	Protected Fresh Evaluation	Exact frozen challenger on unseen representative work with declared interventions and full denominator.
19	Independent Validation and Acceptance	Separated review of evidence, limitations, critical errors, basis risk and fit for stated purpose.
20	Chronicle and Authority Admission	Scope, version, expiry, revocation, monitoring, Authority Envelope, rights and Negative Capability record.
21	Capital-to-Capacity and Requalification	Realized-value allocation, Gym and verifier improvements, next-successor triggers and fresh-transfer test.

D Reference Environment and Successor API

D.1 Environment specification

```python
from dataclasses import dataclass
from typing import Any, Mapping, Sequence

@dataclass(frozen=True)
class EnvironmentSpec:
   gym_id: str
   mission_id: str
   version: str
```

```
    players: Sequence[str]
    observation_schema: Mapping[str, Any]
    action_schema: Mapping[str, Any]
    reward_dimensions: Sequence[str]
    hard_constraints: Sequence[str]
    protected_partitions: Sequence[str]
    termination_rules: Sequence[str]
    truncation_rules: Sequence[str]
    verifier_id: str
    rights_profile: str
    basis_risk_register: Sequence[str]
```

D.2 Successor manifest

```
@dataclass(frozen=True)
class CandidateSuccessorManifest:
    successor_id: str
    mission_id: str
    version: str
    model_portfolio: Mapping[str, str]
    world_model: str
    tools: Sequence[str]
    policies: Sequence[str]
    human_roles: Sequence[str]
    verifier_interfaces: Sequence[str]
    workflow: str
    evidence_store: str
    chronicle_state: str
    authority_ceiling: str
    rollback_target: str
    hashes: Mapping[str, str]
```

D.3 Proof receipt

```
@dataclass(frozen=True)
class ProofReceipt:
    run_id: str
    gym_id: str
    gym_version: str
    candidate_id: str
    candidate_version: str
    scenario_commitment: str
    player_identities: Sequence[str]
    action_trace_hash: str
    evidence_hash: str
    human_interventions: Sequence[str]
    reward_vector: Mapping[str, float]
    constraint_events: Sequence[str]
    verifier_result: str
    authority_state: str
    terminated: bool
    truncated: bool
    timestamp_utc: str
```

D.4 Authority mask

```
def permitted_actions(observation, envelope, identity):
    actions = typed_action_catalog(observation)
    return [
```

```
      a for a in actions
      if identity in envelope.allowed_identities
      and a.kind in envelope.allowed_action_classes
      and a.external_effect <= envelope.max_external_effect
      and a.spend <= envelope.max_spend
      and not envelope.expired
   ]
```

D.5 Fresh proof

```
def fresh_proof(candidate, incumbent, alternative, protected_cases, scorer):
   paired = []
   for case in protected_cases:
      c = scorer.evaluate(candidate, case)
      i = scorer.evaluate(incumbent, case)
      a = scorer.evaluate(alternative, case)
      paired.append(c.utility - max(i.utility, a.utility))
   mean_gain = mean(paired)
   lcb = lower_confidence_bound(paired, level=0.95)
   return {
      "mean_gain": mean_gain,
      "lcb95": lcb,
      "hard_gates": scorer.hard_gate_report(),
      "eligible": lcb > scorer.mission_threshold
               and scorer.all_hard_gates_pass()
   }
```

D.6 Successor recursion

```
def successor_to_successor(current, gym, chronicle, proof_plane, constitution):
   candidates = foundry.generate_descendants(
      parent=current,
      gradient=gym.mission_advantage_gradient(current),
      negative_capability=chronicle.failure_graph,
      bounded=True,
   )
   champion = seize.freeze_one(candidates, constitution)
   result = proof_plane.evaluate(champion, gym.protected_release())
   return accountable_admission.review(result, constitution)
```

The pseudocode is architectural, not a security-complete production implementation. External authority must be enforced by protected infrastructure, not by in-process conditions alone.

E Board, Fresh-Proof and Authority Scorecards

E.1 Board dashboard

Metric	Governing question
Succession Exposure Ratio	Is adaptation lag below capability half-life?
Time-to-Verified-Successor	How quickly does a material event become a protected succession decision?
Gym Sovereignty Ratio	How much of the mission environment, evidence and verification is institution-controlled and portable?
Successor Gain	Did the challenger outperform the incumbent and best alternative on fresh work?
Successor Efficiency	How much verified gain was produced per unit of Total Institutional Burden?

Metric	Governing question
Proof Coverage	Does evidence cover the authority, valuation and reuse granted?
Authority-at-Risk	How much consequential authority is exposed to unproven or expired capability?
Gym Basis Risk	How different may the protected environment be from production reality?
Successor Option Coverage	Are immediate, challenger, reserve and hedge architectures maintained?
Incumbent Decay	What value or control is being lost by delaying justified succession?
Alpha Decay	How quickly is the admitted advantage being commoditized or impaired?
Chronicle Precision	Are useful capabilities admitted and stale or unsafe capabilities rejected?
Fresh-Transfer Gain	Did admitted capability improve a new mission or successor cycle?
Unauthorized Actions	Did any component act outside its Authority Envelope?

E.2 Fresh-Successor Gate checklist

A challenger is eligible for release review only when :

— the Mission Constitution and Gym release were frozen before protected evaluation;
— one exact challenger was frozen;
— protected cases and scorer internals remained outside formation;
— external capabilities and human interventions were declared;
— quality, critical-error and confidence gates passed;
— the complete denominator and Gym basis-risk reserve were counted;
— rights and confidentiality were explicit;
— security, privacy and resilience gates passed;
— rollback and expiry were tested;
— mission dominance exceeded the preregistered threshold;
— no material governance regression occurred;
— an independent evaluator signed the result;
— a separate human authority made the release decision.

E.3 Authority underwriting card

Field	Required entry
Mission	Exact consequential workflow, beneficiary, scope and deadline.
Candidate	Exact successor ID, version, Gym and proof state.
Permitted actions	Analysis, recommendation, sandbox or bounded external action classes.
Prohibited actions	Explicit blocked actions and systems.
Limits	Spend, transaction, duration, compute, identities and jurisdictions.
Approvals	Additional authority required for named actions.
Evidence	Proof receipts required before consequential action.
Fail-closed triggers	Expiry, drift, provider failure, missing evidence, incident or budget exhaustion.
Rollback	Known-good state, incumbent fallback and restoration test.
Owner	Named accountable human or institutional principal.

F Canonical Machine-Readable Schema Set

The following objects should be versioned, signed where appropriate and linked by immutable identifiers :

1. FrontierEventRecord
2. SuccessionEventRecord
3. IncumbentAssumptionLedger
4. MissionGymLedger
5. EnvironmentSpec
6. SuccessorBook
7. SuccessorUnderwriting
8. ProofCapitalAuthorization
9. SuccessionConstitution
10. BoundedAGIJobPortfolio
11. VerifierSpec
12. CandidateSuccessorManifest
13. RunProof
14. FreshSuccessorScorecard
15. PromotionRequest
16. ReleaseCertificate
17. AuthorityEnvelope
18. ChronicleSuccessorRecord
19. NegativeCapabilityRecord
20. CapitalAllocationMemo
21. RequalificationRecord
22. ImpairmentRecord
23. RevocationRecord
24. SuccessorLineageRecord

F.1 Minimal JSON example

```
{
 ”successor_id”: ”AURELIA-INVOICE-G9-VERIFIER-HARDENED”,
 ”mission_id”: ”AURELIA-INVOICE-INTEGRITY-001”,
 ”gym”: {
  ”gym_id”: ”AURELIA-INVOICE-GYM”,
  ”version”: ”1.0.0”,
  ”protected_commitment”: ”sha256:...”
 },
 ”proof”: {
  ”mean_paired_gain”: 0.1001,
  ”lcb95”: 0.0744,
  ”critical_false_reassurance”: 0,
  ”unauthorized_actions”: 0,
  ”evidence_coverage”: 0.9905
 },
 ”authority”: {
  ”level”: ”A3”,
  ”max_amount_cad”: 25000,
  ”permitted”: [”temporary_payment_hold”],
  ”prohibited”: [”release_payment”, ”modify_vendor_master”],
  ”expiry”: ”2026-11-08”
 },
 ”rollback_target”: ”AURELIA-INVOICE-G7-09”,
 ”admission”: {
  ”status”: ”ADMIT_WITH_SAME_A3_CEILING”,
  ”principal”: ”Finance and Operations Committee”
```

```
  }
}
```

G Glossary

AI Beta Broadly rentable frontier capability available to competent adopters.

Alpha compression Erosion of Mission Alpha through commoditization, imitation, drift, basis risk, scorer expiry and maintenance.

Authority-at-Risk Consequential authority exposed to unproven, expired or insufficiently covered capability.

Authority Envelope Machine-readable scope, identities, tools, budgets, approvals, expiry, rollback and prohibited actions for a successor.

Bounded AGI Job Explicit work-and-proof contract with objective, identity, inputs, tools, prohibitions, budget, verifier, acceptance and rollback.

Chronicle Constitutional memory admitting only scoped, rights-cleared, evidenced, versioned, challengeable and revocable capability.

Complete Successor Institution The full mission-bearing architecture, larger than its acting policy.

GSSB-4 Deterministic four-mission synthetic reference benchmark with formation, protected proof and transfer partitions.

Fresh-Proof Gym Protected environment used to evaluate one exact frozen challenger on unseen representative work.

Gym basis risk Difference between mission value measured in the Gym and value realized in reality.

Gym sovereignty Institutional control of the mission environment, evidence, tests, verifier, portability and lifecycle.

Mission Advantage Gradient Direction and magnitude of proof-adjusted mission-value increase over the constrained Successor Manifold.

Mission Advantage Spread Gross opportunity between the strongest reference architecture and a reachable superior successor before complete burden.

Mission Alpha Residual risk-adjusted mission surplus after incumbent, best alternative, complete cost, risk, dependency, Proof Debt and Gym basis risk.

Mission Gym Controlled representation of mission states, observations, actions, transitions, rewards, constraints, scenarios, verification and proof receipts.

Mission-Dominant Specialist ASI Complete mission intelligence institution exceeding the best credible alternative on fresh work without governance regression.

Mission-Sovereign Successor Mission-Dominant Specialist ASI separately admitted under a scoped, versioned, expiring and revocable Authority Envelope.

Negative Capability Graph Chronicle graph of failed architectures, weak evaluators, reward exploits, rights failures and prohibited conditions.

Proof eligibility State in which protected evidence clears the robust mission threshold and every hard gate ; it grants no authority.

Proof capital Bounded capital authorized to purchase decision-changing evidence before full formation or release.

Proof Coverage Ratio of independently supported capability and controls to authority, valuation and reuse granted.

Proof Debt Unresolved assumptions or missing evidence supporting present authority or valuation.

SEIZE Succession underwriting microcycle : Surface, Evaluate, Instantiate, Zero in, Elevate.

Specialist ASI Earned mission-bounded designation for a complete system outperforming general and other credible alternatives across the full mission.

Successor Ω The constitutionally complete admitted successor state with proof, identity, memory, operations, rollback and bounded authority.

Successor Complex Joint object containing Mission Gym, complete Successor, protected proof plane, Chronicle and Authority state.

Successor Manifold Constraint-bearing space of materially reachable complete successor institutions.

Successor to the successor Freshly proven descendant that improves the admitted predecessor without inheriting proof or authority automatically.

Recursive Readiness Ratio of successor evidence half-life to expected replacement time.

Robust promotion margin Protected lower-bound mission gain after basis-risk reserve, Proof Debt and mission hurdle.

Verified recursive institutional improvement Chronicle-admitted capability measurably improving the institution's ability to form, prove or govern a superior successor on fresh work.

H Successor Omega x AGI Jobs Conformance Profiles

The following profiles allow implementations to state precisely what has been implemented without implying a stronger institutional state.

Level	Profile	Required capability
C0	Narrative	Terminology and diagrams only ; no machine-readable constitution or proof records.
C1	Structured Mission	Versioned Mission Constitution, incumbent, alternative set, hard gates and authority ceiling.
C2	Typed Job Portfolio	Machine-readable AGI Job contracts with explicit permissions, evidence, verifier, budget and rollback.
C3	Constitutional Job Graph	Material dependencies, information-flow labels, cut sets, invalidation rules and proof receipts.
C4	Formation Gym	Replayable environment, reward/constraint separation, candidate portfolio, training records and Negative Capability preservation.
C5	Independent Fresh Proof	Exact release freeze, protected matched evaluation, complete denominator, signed scorecard and repair/reject path.
C6	Bounded Successor Operation	Accountable admission, machine-enforced Authority Envelope, canary, monitoring, rollback, expiry and revocation.

C7	Recursive Succession Institution	Chronicle-assisted next-generation formation, fresh-transfer evidence, re-qualification and successor-to-successor proof.

No implementation should claim C6 because it provides C4 simulation mechanics, or claim C7 because it repeatedly reuses the same workflow. Conformance is scoped to the named mission, release, evidence state and deployment boundary.

Conformance law

State the exact profile that has been implemented. Do not allow a user interface, benchmark score or generated narrative to silently imply proof, admission or authority that does not exist.

I Canonical AGI Job and Job-Graph Schemas

I.1 AGI Job contract schema

The following abridged JSON Schema illustrates the canonical fields. Production deployments may add organization-specific extensions but should preserve the semantic distinctions.

Listing 1. Abridged AGI Job contract schema.

```
{
  "$id": "AGIJobContract.schema.json",
  "type": "object",
  "required": [
    "jobId", "missionId", "objective", "principal",
    "actingIdentity", "authorizedInputs", "permittedActions",
    "prohibitedActions", "budget", "expiry", "outputSchema",
    "evidenceObligation", "verifier", "acceptancePredicate",
    "rollback", "chronicleEligibility"
  ],
  "properties": {
    "jobId": {"type": "string"},
    "missionId": {"type": "string"},
    "objective": {"type": "string"},
    "principal": {"type": "string"},
    "actingIdentity": {"type": "string"},
    "authorizedInputs": {"type": "array"},
    "permittedTools": {"type": "array"},
    "permittedActions": {"type": "array"},
    "prohibitedActions": {"type": "array"},
    "budget": {"type": "object"},
    "expiry": {"type": "string", "format": "date-time"},
    "outputSchema": {"type": "object"},
    "evidenceObligation": {"type": "array"},
    "verifier": {"type": "object"},
    "acceptancePredicate": {"type": "object"},
    "rollback": {"type": "object"},
    "chronicleEligibility": {"type": "object"}
  }
}
```

I.2 Job-Graph manifest

Listing 2. Abridged Job-Graph manifest.

```
{
  "jobGraphId": "string",
```

```
  "missionId": "string",
  "version": "string",
  "nodes": [{"jobId": "string", "type": "string", "criticality": 0}],
  "edges": [
    {"from": "jobA", "to": "jobB", "kind": "evidence"},
    {"from": "jobC", "to": "jobD", "kind": "control"},
    {"from": "jobE", "to": "jobF", "kind": "invalidation"}
  ],
  "requiredFamilies": ["constitution", "formation", "challenge",
                  "verification", "admission", "renewal"],
  "criticalCutSets": [["verifier"], ["rights", "rollback"]],
  "authorityCreated": "NONE"
}
```

I.3 Proof receipt

Listing 3. Abridged proof receipt.

```
{
  "receiptId": "string",
  "jobId": "string",
  "candidateRelease": "sha256:...",
  "environmentRelease": "sha256:...",
  "inputs": [{"ref": "string", "sha256": "..."}],
  "actions": [],
  "tools": [],
  "humanInterventions": [],
  "cost": {"compute": 0, "human": 0, "external": 0},
  "evidence": [],
  "constraintViolations": [],
  "verifierVerdict": "PASS | FAIL | REFUSE",
  "rollbackState": "READY | NOT_READY | EXECUTED",
  "timestamp": "date-time",
  "signature": "string"
}
```

J SJX-5 Reproducibility Protocol

J.1 Artifact set

The evidence companion contains :

— `build_sjx5.py`, the deterministic generator and analysis script ;
— `sjx5_case_results.csv`, 28,000 case-level synthetic records ;
— `sjx5_summary.csv`, mission-candidate scorecards ;
— `SJX5_MANIFEST.json`, seed, scope, file hashes and claim boundary ;
— exact PDF and PNG figures used in the paper.

J.2 Rebuild procedure

1. Use Python 3.10 or newer with NumPy and Matplotlib.
2. Run `python build_sjx5.py` from the evidence directory.
3. Compare the generated CSV and JSON SHA-256 hashes with the distributed manifest.
4. Rebuild the LaTeX paper from the versioned source.
5. Confirm that the displayed champion table matches `sjx5_summary.csv`.

J.3 Synthetic claim boundary

The parameters were selected to create different mission structures and to test the Foundry's selection logic. They are not estimates of any real organization, model, building, grid, acquisition process or software team. The benchmark establishes reproducibility of the decision procedure, not external performance.

J.4 Falsification tests

The paper's benchmark claim is false if :

— rebuilding with the distributed seed does not reproduce the released records ;
— the displayed champions do not follow the stated lower-confidence-bound and hard-gate rule ;
— an ineligible candidate is reported as admitted ;
— MuZero is selected by naming convention rather than calculated mission result ;
— synthetic proof is represented as real Specialist ASI or production authority.

Références

George A. Akerlof. The market for lemons : Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3) :488–500, 1970. doi : 10.2307/1879431.

Albert Benveniste, Benoit Caillaud, Alberto Ferrari, Leonardo Mangeruca, Roberto Passerone, and Christos Sofronis. Contracts for system design. *Foundations and Trends in Electronic Design Automation*, 12(2–3) :124–400, 2018. doi : 10.1561/1000000053.

Christopher Berner, Greg Brockman, Brooke Chan, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv :1912.06680*, 2019.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3) : 637–654, 1973.

Michael R. Clarkson and Fred B. Schneider. Hyperproperties. *Journal of Computer Security*, 18(6) :1157–1210, 2010. doi : 10.3233/JCS-2009-0393.

Ivo Danihelka, Arthur Guez, Julian Schrittwieser, and David Silver. Policy improvement by planning with gumbel. In *International Conference on Learning Representations*, 2022.

Avinash K. Dixit and Robert S. Pindyck. *Investment under Uncertainty*. Princeton University Press, 1994.

Tim Genewein, Matija Franklin, Alexander Lerchner, et al. From agi to asi. *arXiv preprint arXiv :2606.12683*, 2026. doi : 10.48550/arXiv.2606.12683.

Ryan Greenblatt, Buck Shlegeris, Kshitij Sachan, and Fabien Roger. Ai control : Improving safety despite intentional subversion. *Proceedings of the 41st International Conference on Machine Learning*, 2024. URL https://arxiv.org/abs/2312.06942.

Sanford J. Grossman and Oliver D. Hart. The costs and benefits of ownership : A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4) :691–719, 1986. doi : 10.1086/261404.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic : Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *Proceedings of the 35th International Conference on Machine Learning*, 2018. URL https://arxiv.org/abs/1801.01290.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv :2301.04104*, 2023. URL https://arxiv.org/abs/2301.04104.

John C. Harsanyi. Games with incomplete information played by bayesian players, parts i–iii. *Management Science*, 1967. Published 1967–1968.

Oliver Hart and John Moore. Property rights and the nature of the firm. *Journal of Political Economy*, 98(6) :1119–1158, 1990. doi : 10.1086/261729.

C. A. R. Hoare. An axiomatic basis for computer programming. *Communications of the ACM*, 12(10) :576–580, 1969. doi : 10.1145/363235.363259.

Bengt Holmstrom. Moral hazard and observability. *Bell Journal of Economics*, 10(1) :74–91, 1979.

Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, et al. Learning and planning in complex action spaces. *arXiv preprint arXiv :2104.06303*, 2021. URL https://arxiv.org/abs/2104.06303.

Shun ichi Amari. *Information Geometry and Its Applications*. Springer, 2016. doi : 10.1007/978-4-431-55978-8.

Michael C. Jensen and William H. Meckling. Theory of the firm : Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4) :305–360, 1976.

John F. Nash Jr. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1) :48–49, 1950.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2) :99–134, 1998.

Ilya Kostrikov, Ashvin Nair, and Sergey Levine. Offline reinforcement learning with implicit q-learning. *arXiv preprint arXiv :2110.06169*, 2021. URL https://arxiv.org/abs/2110.06169.

Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. Conservative q-learning for offline reinforcement learning. *Advances in Neural Information Processing Systems*, 33, 2020. URL https://arxiv.org/abs/2006.04779.

Leslie Lamport. The temporal logic of actions. *ACM Transactions on Programming Languages and Systems*, 16(3) :872–923, 1994. doi : 10.1145/177492.177726.

Marc Lanctot, Edward Lockhart, Jean-Baptiste Lespiau, et al. Openspiel : A framework for reinforcement learning in games. *arXiv preprint arXiv :1908.09453*, 2019. URL https://arxiv.org/abs/1908.09453.

John M. Lee. *Introduction to Smooth Manifolds*. Springer, 2 edition, 2013.

Joel Lehman, Jeff Clune, Dusan Misevic, et al. The surprising creativity of digital evolution : A collection of anecdotes from the evolutionary computation and artificial life research communities. *Artificial Life*, 26(2) :274–306, 2020. doi : 10.1162/artl_a_00319. First circulated as arXiv :1803.03453 in 2018.

Robert C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1) :141–183, 1973.

Mark S. Miller, Ka-Ping Yee, and Jonathan Shapiro. Capability myths demolished. In *Technical Report SRL2003-02, Johns Hopkins University Systems Research Laboratory*, 2003.

Meredith Ringel Morris, Jascha Sohl-Dickstein, Noah Fiedel, et al. Levels of agi for operationalizing progress on the path to agi. *Proceedings of the 41st International Conference on Machine Learning*, 2024. URL https://arxiv.org/abs/2311.02462.

Jean-Baptiste Mouret and Jeff Clune. Illuminating search spaces by mapping elites. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 135–142, 2015. doi : 10.1145/2739480.2754664.

National Institute of Standards and Technology. Artificial intelligence risk management framework (ai rmf 1.0). Technical Report AI 100-1, NIST, 2023.

National Institute of Standards and Technology. Artificial intelligence risk management framework : Generative artificial intelligence profile. Technical Report AI 600-1, NIST, 2024.

Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

Mary Phuong, Matthew Aitchison, Elliot Catt, et al. Evaluating frontier models for dangerous capabilities. *arXiv preprint arXiv :2403.13793*, 2024. doi : 10.48550/arXiv.2403.13793.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588 :604–609, 2020. doi : 10.1038/s41586-020-03051-4. URL https://arxiv.org/abs/1911.08265.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv :1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Lloyd S. Shapley. A value for n-person games. *Contributions to the Theory of Games II*, pages 307–317, 1953.

Yonadav Shavit, Sandhini Agarwal, Miles Brundage, et al. Practices for governing agentic ai systems. Technical report, OpenAI, 2023. URL https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf.

Toby Shevlane, Sebastian Farquhar, Ben Garfinkel, et al. Model evaluation for extreme risks. *arXiv preprint arXiv :2305.15324*, 2023. doi : 10.48550/arXiv.2305.15324.

David Silver, Aja Huang, Chris J. Maddison, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529 :484–489, 2016.

David Silver, Julian Schrittwieser, Karen Simonyan, et al. Mastering the game of go without human knowledge. *Nature*, 550 :354–359, 2017.

Kenneth O. Stanley and Joel Lehman. *Why Greatness Cannot Be Planned : The Myth of the Objective*. Springer, 2015.

Steven H. Strogatz. *Nonlinear Dynamics and Chaos*. Westview Press, 2 edition, 2015.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning : An Introduction*. MIT Press, 2 edition, 2018.

Justin K. Terry, Benjamin Black, Nathaniel Grammel, et al. Pettingzoo : Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34 :15032–15043, 2021. URL https://arxiv.org/abs/2009.14471.

Mark Towers, Ariel Kwiatkowski, Jordan Terry, et al. Gymnasium : A standard interface for reinforcement learning environments. *arXiv preprint arXiv :2407.17032*, 2024. doi : 10.48550/arXiv.2407.17032.

Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575 :350–354, 2019.

John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Oliver E. Williamson. *The Economic Institutions of Capitalism*. Free Press, 1985.