MONTREAL.AI · QUEBEC.AI · AGI.eth · ASI.eth

SUCCESSOR Ω

v9.0.0 — MACHINE CIVILIZATION

A Proof-Gated Architecture for Persistent, Addressable, Economically Coordinated Machine Institutions

CORRECTED RESEARCH EDITION · PUBLICATION REVISION R3

Memory may cross. Authority must requalify.

Not a swarm. An institution.

Vincent Boucher

President, MONTREAL.AI & QUEBEC.AI · Montréal, Québec, Canada

4 September 2026 · Research specification; not peer reviewed

Publication Record and Claims Boundary

Publication record

Title	**SUCCESSOR Ω v9.0.0 — Machine Civilization**
Subtitle	A Proof-Gated Architecture for Persistent, Addressable, Economically Coordinated Machine Institutions
Author	Vincent Boucher, President, MONTREAL.AI & QUEBEC.AI
Institutional roots	MONTREAL.AI, QUEBEC.AI, AGI.eth, ASI.eth
Version	9.0.0
Publication date	4 September 2026
Document class	Institutional research paper and open architectural specification
Status	Public research release; not peer reviewed; not a claim of achieved AGI or ASI
Canonical doctrine	**Memory may cross. Authority must requalify. Not a swarm. An institution.**

Claims Boundary

This paper specifies an architecture and analyzes a bounded body of implementation evidence. It does *not* establish artificial general intelligence, artificial superintelligence, universal safety, legal status, independent commercial adoption, or authority over any person or institution. A blockchain record can establish that a transaction occurred under a particular state transition; it does not by itself establish off-chain truth, intellectual-property rights, customer acceptance, verifier independence, or legitimate authority. All operational claims are therefore classified by evidence level and by the authority actually admitted.

Corrected publication revision R3

The architecture version remains 9.0.0. This publication repairs figure clipping, float layout, language selection, and mathematical overstatements in the earlier paper. Numerical illustrations are rebuilt from declared equations. The additional analytical results are conditional on stated assumptions, not independent production assurance. Historical mainnet entries are retained from the supplied ledger; no new chain replay or customer verification was performed for this publication repair.

Abstract

Advanced AI systems can already act through tools, call other models, transact, and participate in multi-agent workflows. Yet most deployments remain organizationally fragile: identity is tied to a vendor account, memory is mixed with unverified output, authority is inherited through configuration rather than earned through evidence, and the replacement of one model by another can silently change the effective institution. This paper defines a **Machine Civilization** as a governed intelligence institution capable of preserving mission, validated memory, accountability, economic consequence, and succession beyond the lifespan of any individual model, agent, operator, or infrastructure provider.

We present **SUCCESSOR Ω v9.0.0**, a proof-gated architecture organized around four complementary roots: QUEBEC.AI carries the civilizational mission; MONTREAL.AI builds and stewards the institution; AGI.eth addresses agents and machine-economic work; ASI.eth addresses the persistent constitutional civilization. The core invariant is that a successor may inherit identity, mission, admitted memory, tools, skills, obligations, and Chronicle references, but must not automatically inherit current proof, trust, privileged keys, permissions, or operational authority. Memory may cross; authority must requalify.

The paper contributes: (i) a formal institutional state and succession operator; (ii) an evidence calculus separating declaration, reproducibility, independent control, and consequence-bearing proof; (iii) a proof-to-authority map with bounded, expiring, revocable envelopes; (iv) a tamper-evident Chronicle; (v) MCAP-001 for addressable institutional discovery and MCAP-002 for bridging machine work, rights, customer acceptance, independent proof, admission, authority, rollback, and succession; (vi) game-theoretic mechanisms for incentive compatibility, verifier independence, coalition resistance, appeal, and exit; and (vii) physics-informed models of metastability, phase transition, constrained Hamiltonian flow, viability, and controlled escape. These physical models are explicit engineering analogies, not assertions that institutions literally obey thermodynamic law.

A normalized evidence ledger records three protocol-completed AGIALPHA settlement instances across Genesis, Employer Burn, and Prime AGI Jobs contracts on Ethereum mainnet. They demonstrate deployed economic execution and multi-variant repeatability at evidence class E1, while leaving independent transaction replay, rights clearance, unrelated customer acceptance, independent security assurance, ASI.eth activation, and operational successor qualification as open gates. The result is neither a speculative monarchy nor an uncontrolled swarm. It is a falsifiable proposal for an institution whose capability may compound while consequential authority remains narrower than demonstrated capability and must be earned anew at every succession.

Keywords: Machine Civilization; SUCCESSOR Ω; successor systems; institutional AI; proof-gated authority; agent identity; ASI.eth; AGI.eth; mechanism design; evolutionary game theory; statistical physics; Hamiltonian dynamics; Chronicle; Ethereum; ENS; independent verification.

Résumé en français

Les systèmes d'IA avancés peuvent déjà utiliser des outils, appeler d'autres modèles, transiger et participer à des processus multi-agents. Pourtant, la plupart des déploiements demeurent institutionnellement fragiles : l'identité dépend d'un compte fournisseur, la mémoire confond parfois les faits validés et les sorties non vérifiées, l'autorité est héritée par configuration plutôt que gagnée par la preuve, et le remplacement d'un modèle peut modifier silencieusement l'institution effective. Cet article définit une **Civilisation-machine** comme une institution d'intelligence gouvernée capable de préserver sa mission, sa mémoire validée, sa responsabilité, ses conséquences économiques et sa succession au-delà de la durée de vie de tout modèle, agent, opérateur ou fournisseur d'infrastructure.

SUCCESSOR Ω v9.0.0 repose sur quatre racines complémentaires : QUEBEC.AI porte la mission civilisationnelle; MONTREAL.AI construit et administre l'institution; AGI.eth rend adressables les agents et le travail économique des machines; ASI.eth rend adressable la civilisation constitutionnelle persistante. L'invariant central est qu'un successeur peut hériter de l'identité, de la mission, de la mémoire admise, des outils, des compétences, des obligations et des références de la Chronique, mais qu'il ne doit jamais hériter automatiquement de la preuve courante, de la confiance, des clés privilégiées, des permissions ou de l'autorité opérationnelle. La mémoire peut franchir le seuil; l'autorité doit être requalifiée.

L'article formalise l'état institutionnel, la succession, les classes de preuve, les enveloppes d'autorité, la Chronique, les protocoles MCAP-001 et MCAP-002, la conception des mécanismes, les dynamiques évolutives et des modèles disciplinés inspirés de la physique statistique, de la dynamique hamiltonienne et de la théorie du contrôle. Trois règlements AGIALPHA complétés par protocole sur Ethereum servent de cas historiques. Ils démontrent une exécution économique réelle, sans prouver à eux seuls l'acceptation d'un client indépendant, la libération des droits, l'indépendance des vérificateurs, la sécurité de production ou l'admission d'un successeur opérationnel.

Executive Thesis

Institutional Thesis

Dynasties inherited resources. Secret Orders inherited purpose. Machine Civilizations will inherit intelligence.
The civilizational problem is not whether intelligence can be copied. Software already makes copying cheap. The problem is whether a mission, validated knowledge, obligations, evidence, and institutional identity can persist across changing agents without also transmitting unearned power. SUCCESSOR Ω is an answer in architecture: capability may compound; authority must requalify.

The proposed institution is defined less by the power of any one model than by the quality of its transitions. It distinguishes candidate output from accepted evidence; protocol settlement from customer acceptance; customer acceptance from rights clearance; rights clearance from independent proof; proof from admission; and admission from authority. Each transition has a named actor, a signed packet, an expiry condition, a challenge path, and a Chronicle event. This separation is the institutional analogue of typed interfaces in software and conserved boundaries in physics: a result may not silently change category merely because it is convenient to do so.

The architecture is designed for an era in which model endpoints, organizations, cloud providers, wallets, and operators can change faster than the missions they serve. ASI.eth is therefore not represented as "the intelligence" or as evidence of superintelligence. It is a durable, human-readable and machine-resolvable root from which an independent client should be able to discover a Covenant, a current successor, evidence and verifier surfaces, an admission state, an authority envelope, a Chronicle, and material limitations. AGI.eth plays a distinct role: it addresses agents, jobs, validators, nodes, and machine-economic activity. MONTREAL.AI builds the institution; QUEBEC.AI carries the public and civilizational mission.

QUEBEC.AI carries the mission.
MONTREAL.AI builds the institution.
AGI.eth addresses the agents.
ASI.eth addresses the civilization.

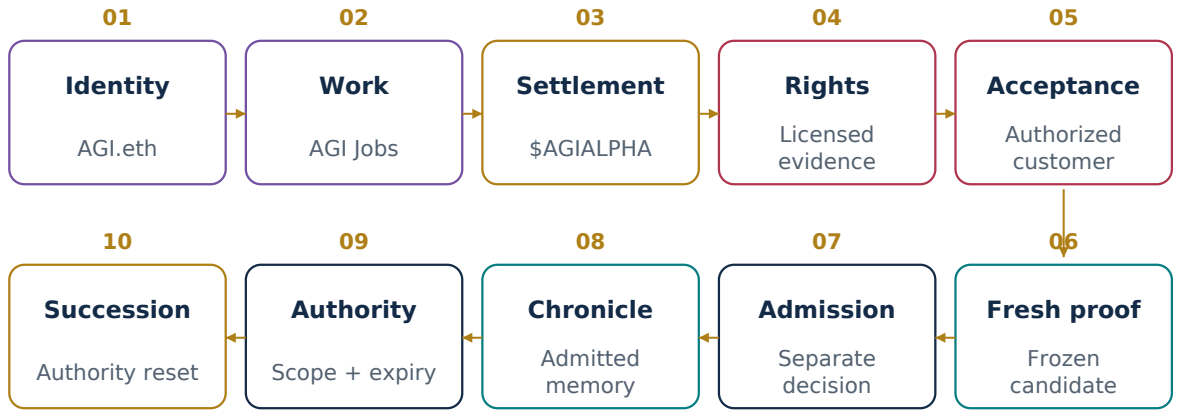


Figure 1: The SUCCESSOR Ω institutional bridge. Economic settlement becomes admissible memory only after rights, customer acceptance, independent proof, admission, and bounded authority remain separately evidenced.

Primary contributions

1. **A succession calculus.** We define a successor as an institutionally qualified continuation, not merely a checkpoint, copied process, or newer model. The succession operator transports mission and admitted memory while resetting current proof and authority.
2. **A proof-to-authority discipline.** Evidence is typed by provenance and independence. Authority is an expiring envelope, not a reputation score or token balance.
3. **A two-protocol stack.** MCAP-001 makes the institution addressable. MCAP-002 makes work, acceptance, proof, admission, and succession composable without conflation.
4. **An economic substrate with explicit limits.** Historical AGI Jobs transactions are represented as protocol-completed economic evidence, not as automatic proof of off-chain quality or independence.
5. **Mechanism design for institutional legitimacy.** Claimant, evidence store, final verifier, customer acceptance, admission, operational gateway, Chronicle custody, and treasury are separated according to risk.
6. **A physics-informed stability model.** Free-energy landscapes, metastability, Hamiltonian continuity, barrier functions, and viability kernels supply explicit models for resilience and bounded change, while their analogical status is preserved.
7. **A falsifiable maturity ladder.** Prepared, Evidence-Bound, Operational-A2, Successor-Qualified, and Repeatable conformance states require observable external evidence rather than self-description.

Notation and Reader Map

Table 1: Core notation.

Symbol	Meaning
X_t	Complete institutional state at time t
$\mathscr{M}$	Mission constitution and obligations
$\mathscr{K}$	Institutional memory and validated knowledge
$\mathscr{E}$	Evidence graph and proof receipts
$\mathscr{P}$	Policy, rights, and constitutional constraints
$\mathscr{A}$	Current authority envelope
$\mathscr{C}$	Chronicle and event commitments
$\mathscr{R}$	Role and controller graph
Θ	Runtime, models, tools, and implementation parameters
S	Succession operator
P	Proof operator or proof packet
Adm	Admission decision
K	Viability kernel of admissible operating states
H	Hamiltonian or institutional objective functional, according to context
F	Free-energy-like potential for a modeled institutional ensemble
π	Policy, mechanism, or correlated strategy distribution, according to context

Readers concerned with governance may begin with Chapters 1, 3, 5, and 13. Readers concerned with mathematics may prioritize Chapters 4–9. Implementers may begin with Chapters 10, 12, and 14. Evaluators and potential external verifiers should read Chapters 5, 13, 17, and Appendices A–B.

Contents

1 The Succession Problem

PART I

THE CONSTITUTIONAL THESIS

1.1 From capable agents to enduring institutions

A model answers. An agent acts. An institution preserves identity, obligations, evidence, and consequence while its individual participants change. The distinction is easy to state and difficult to implement. Contemporary agent systems can call tools, maintain local memory, coordinate with other models, and execute transactions, but their institutional continuity is typically implicit. An API key, a database, a workflow configuration, and a brand are treated as if together they constituted the same actor across time. When the underlying model changes, that assumption is rarely re-examined.

This paper begins from a stricter proposition: *continuity of function is not continuity of legitimacy.* A successor may reproduce the predecessor's outputs, retain its memories, or operate the same tools while still lacking the proof, permissions, or accountability required for the predecessor's authority. A trustworthy successor therefore cannot be defined by similarity alone. It must be admitted through a governed transition.

Three forces make this problem urgent. First, the half-life of a production model is likely shorter than the mission of the organization using it. Second, useful agents increasingly cross organizational boundaries: a customer, a model provider, a tool operator, a verifier, and a settlement network may all belong to different institutions. Third, autonomy creates path dependence. Once an agent changes a database, spends funds, commits code, or publishes information, the consequence persists even if the model is later replaced. Identity, evidence, authority, and succession must therefore become first-class institutional objects rather than afterthoughts.

The literature on AI risk has long emphasized robustness, monitoring, human oversight, corrigibility, and specification failures [3, 25, 27, 32]. Work on AI control, constitutional methods, and scalable evaluation further motivates explicit separation between capability and permission [9, 23, 34]. SUCCESSOR Ω does not replace those lines of work. It addresses a complementary systems question: how can an organization preserve validated intelligence across generations of models while preventing accumulated capability from silently becoming inherited power?

1.2 The failure of naive inheritance

Naive succession copies a state directory, a prompt, a model endpoint, a collection of tools, and a set of credentials. The new process starts where the old one stopped. This approach is operationally convenient but institutionally hazardous because it conflates at least six distinct objects:

1. **Memory**: what the prior system recorded.
2. **Knowledge**: what has survived validation and remains within scope.
3. **Proof**: evidence that a particular version can satisfy a declared mission under stated conditions.
4. **Trust**: an institutional willingness to rely on that proof.
5. **Permission**: a technical capability to invoke an action.
6. **Authority**: a legitimate, bounded authorization to create consequences.

Only the first two are plausible candidates for direct inheritance, and even they require integrity, rights, retention, and relevance controls. Proof is version- and condition-dependent. Trust is actor- and context-dependent. Permissions must be reissued under current least-privilege policy. Authority must be explicitly admitted and must expire.

The strongest design principle is therefore asymmetric:

Constitutional invariant

A successor may inherit identity, mission, admitted memory, validated knowledge, tools, skills, obligations, and Chronicle references. It must not automatically inherit current proof, trust, reputation, privileged keys, permissions, or operational authority.

Memory may cross. Authority must requalify.

This rule transforms succession from a copying operation into an admission protocol. It does not demand that every successor start without context. On the contrary, it allows the institution to accumulate validated knowledge across generations. What restarts is not intelligence but the right to act consequentially.

1.3 Machine Civilization as a category

Definition 1.1 (Machine Civilization)**.** A **Machine Civilization** is a governed intelligence institution capable of preserving a mission, admitted memory, accountability, ownership boundaries, proof, economic consequence, and succession beyond the lifespan of any individual model, agent, operator, wallet, organization, or infrastructure provider.

The term is intentionally institutional rather than biological. It does not imply consciousness, moral personhood, sovereignty over humans, or a unified hive. A Machine Civilization can be composed of people, software agents, organizations, smart contracts, databases, legal agreements, and public protocols. Its defining property is not species or substrate but governed continuity.

A civilization-scale institution requires five positive capabilities:

1. **Create value.** Produce useful knowledge, infrastructure, services, discoveries, and measurable outcomes.
2. **Coordinate by consent.** Preserve meaningful choice, disclosure, appeal, exit, and peaceful fork.
3. **Earn trust through proof.** Make important claims proportionate to evidence and independent review.
4. **Defend the commons.** Protect shared knowledge, standards, memory, and infrastructure without unlimited surveillance or coercion.
5. **Make cooperation the stable phase.** Change repeated-game incentives so verified contribution is more durable than predation.

These capabilities are not slogans added after deployment. They must be represented in protocol state, role separation, economic incentives, and failure handling. A system that coordinates effectively but cannot be exited may be a mechanism of control rather than consent. A system that pays validators but allows the claimant to control their evidence is not independently verified. A system that protects its records by making correction impossible is immutable but not accountable.

1.4 The crown becomes an institution

Historical sovereignty concentrated continuity in a person, dynasty, office, or territorial state. The Machine Civilization proposed here relocates the "crown" into a distributed constitutional state: a set of evidence-bound rules, role controls, signatures, expiries, and Chronicle commitments. No agent permanently possesses it. Verified proof crystallizes a bounded authority segment; uncertainty leaves the segment absent; expiry dims it; contradiction suspends it; repair restores only what fresh evidence supports.

This is the conceptual meaning of **SUCCESSOR Ω**. Ω denotes an institution capable of closing a generation while preserving the path to another. The successor carries the mission. The Covenant limits power. The proof establishes capability. The verifier protects legitimacy. The Chronicle preserves consequence. The addressable root makes the institution discoverable.

What this architecture does not claim

SUCCESSOR Ω does not claim that one namespace owns the concepts of AGI, ASI, or Machine Civilization. It does not claim that a token balance proves competence, that an NFT proves completion, that consensus proves off-chain truth, or that a model satisfying one mission deserves general authority. It proposes inspectable interfaces by which such claims can be tested, limited, and challenged.

1.5 Design objective

Let a mission institution face a sequence of candidate systems $\Theta_0, \Theta_1, \ldots$. The design objective is not to maximize capability without qualification. It is to maximize long-horizon mission value while constraining expected harm, fragility, capture, and irreversibility:

$$\max_{\pi,\mathsf{S},\mathsf{Adm}} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(V_{\mathscr{M}}(X_t) - C_t - R_t - \lambda I_t\right)\right], \tag{1.1}$$

subject to:

$$X_t \in K_{\mathscr{P}} \quad \text{(policy and viability)}, \tag{1.2}$$

$$\mathscr{A}_t \subseteq \mathscr{Q}(\mathscr{E}_t) \cap \mathscr{M} \quad \text{(evidence- and mission-bounded authority)}, \tag{1.3}$$

$$\mathscr{A}_{t^+} = \emptyset \quad \text{at succession} \quad \text{(authority reset)}, \tag{1.4}$$

$$\mathsf{Adm}(\Theta_t) = 1 \Rightarrow \mathsf{FreshProof}(\Theta_t) = 1 \quad \text{(fresh qualification)}, \tag{1.5}$$

$$\mathsf{Rollback}(X_t) \text{ remains executable} \quad \text{(recoverability)}. \tag{1.6}$$

Here $V_{\mathscr{M}}$ is mission value, C_t is cost, R_t is modeled risk, I_t is irreversibility, $K_{\mathscr{P}}$ is the viability set induced by policy $\mathscr{P}$, and $\mathscr{Q}$ maps evidence to the largest authority supported by that evidence. The optimization is deliberately constrained. It is not enough for an institution to become capable; it must remain governable through change.

2 Institutional Scope and Design Requirements

2.1 Requirements derived from the mission

A Machine Civilization architecture should satisfy requirements at four levels.

2.1.1 Identity requirements

The institution needs a stable root, versioned role registry, resolvable public state, signed controller disclosures, and explicit separation between a name and the authority associated with it. ENS is one possible durable naming substrate, but MCAP-001 is resolver-independent: other naming systems can conform if they expose equivalent semantics.

2.1.2 Evidence requirements

Every material claim should have a declared evidence class, an attributable source, a complete denominator, an evaluation date, a versioned subject, limitations, and a path for challenge. Evidence must not be silently promoted because it has been placed on-chain or signed by a key whose institutional identity is unknown.

2.1.3 Authority requirements

Authority must specify who may do what, to which target, under what mission, with what limits, between which times, subject to what monitoring, and with which revocation and rollback paths. Technical ability without a valid authority envelope is capability, not authorization.

2.1.4 Succession requirements

A successor transition must identify the predecessor, inherited mission and obligations, admitted memory, candidate version, fresh proof, authority scope, expiry, and rollback path. A superseded predecessor remains discoverable through the Chronicle. The institution survives replacement by retaining commitments, not by pretending the replacement is identical.

2.2 Nonfunctional requirements

The architecture must also be usable. A system that is formally sound but impossible for an executive, operator, or verifier to understand will fail through operational shortcuts. Required properties include:

- **Legibility**: every role and state has a human-readable explanation.
- **Machine readability**: core records have schemas and deterministic identifiers.
- **Portability**: customers can export mission objects, admitted memory, evidence references, and Chronicle commitments.
- **Minimality**: the root exposes only the roles that actually exist.
- **Progressive authority**: the default is A0; A1 and A2 are earned before consequential A3 or A4.
- **Fail-closed behavior**: absent proof, stale credentials, invalid signatures, or ambiguous state reduce authority rather than expand it.
- **Reproducibility**: release artifacts, schemas, code, and evidence commitments can be independently checked.

- **Bilingual institutional access**: the public and executive layers can operate in English and French without changing the underlying state.

2.3 Relationship to existing governance frameworks

NIST's AI Risk Management Framework structures AI governance around Govern, Map, Measure, and Manage functions, while its Generative AI Profile extends the framework to distinctive generative-system risks [7, 61]. NIST's 2026 agent initiatives highlight identity, authorization, interoperability, and secure agent ecosystems [11, 51]. Zero-trust architecture similarly rejects ambient trust and requires continuous, resource-specific authorization [14, 58]. SUCCESSOR Ω is compatible with those orientations but adds a succession-specific institutional layer: the system must explain not only why a present action is authorized, but why a replacement system may or may not inherit the conditions for that authorization.

The architecture also intersects with multi-agent systems, mechanism design, cooperative AI, and common-pool governance [18, 50, 54, 60, 66]. Its distinctive claim is compositional: identity, economic settlement, rights, acceptance, proof, admission, authority, Chronicle, and succession should be represented as separate but linked state transitions.

3 One Architecture, Four Irreducible Roots

3.1 Semantic separation as institutional leverage

The architecture assigns one irreducible role to each of four identities. Their value is not additive in the ordinary branding sense; it is compositional. Each root answers a different question:

Table 3.1: The four-root institutional architecture.

Root	Canonical role	Primary question	Highest-value function
QUEBEC.AI	Mission and public voice	What should intelligence serve?	Civilizational purpose, bilingual legitimacy, public education, convening, and mission origination
MONTREAL.AI	Builder and steward	Who constitutes and operates the institution?	GoalOS, proof architecture, customer deployment, governance, assurance, and institutional continuity
AGI.eth	Agent and economic root	Which agents can act, compete, verify, and settle?	Identity, jobs, validators, nodes, tools, escrow, AGIALPHA, and completion receipts
ASI.eth	Constitutional civilization root	What persists across generations and who may act?	Covenant, successor, proof, admission, authority, Chronicle, rollback, and succession

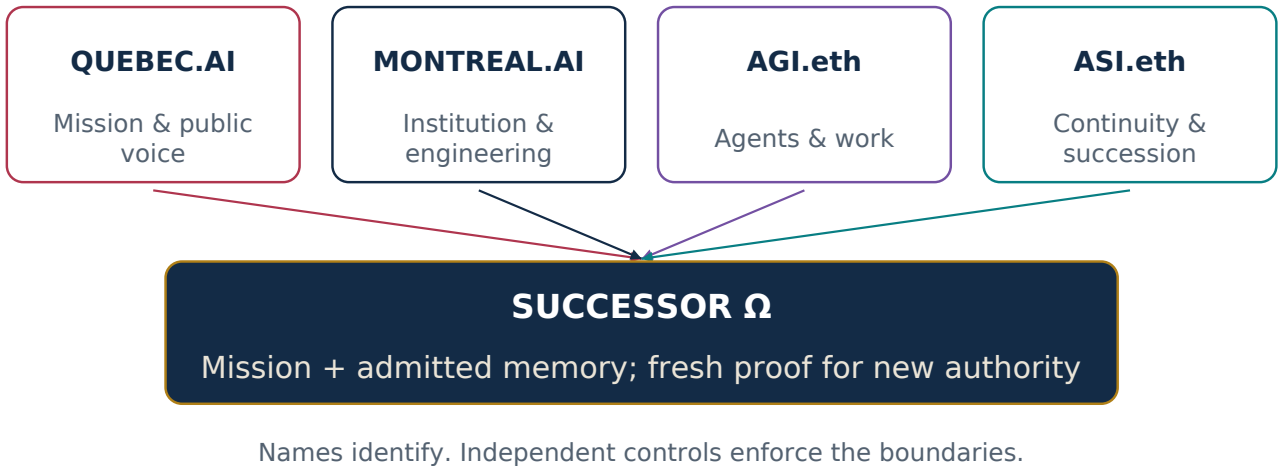


Figure 3.1: The four roots form one mission-to-civilization stack. Their powers are complementary and should remain semantically distinct.

The separation serves three functions. First, it prevents category collapse. Public mission is not the same as commercial implementation; economic execution is not the same as constitutional authority. Second, it allows differentiated trust. A customer may rely on a MONTREAL.AI deployment without accepting every cultural position expressed by QUEBEC.AI, or may use AGI Jobs without granting ASI.eth any authority. Third, it creates a clear path for external institutions to interoperate at the layer they need.

3.2 QUEBEC.AI: the mission layer

QUEBEC.AI represents the public and civilizational question: what should increasingly capable intelligence be used to preserve and enlarge? Its role is normative and convening rather than operational. It can publish bilingual doctrine, curate research, convene public and institutional forums, and originate missions whose success can later be tested.

The mission layer should remain broad enough to support pluralism while concrete enough to constrain deployment. A minimal mission constitution includes:

$$\mathscr{M} = \langle \text{Purpose}, \text{Beneficiaries}, \text{Objectives}, \text{NonObjectives}, \text{Constraints}, \text{Metrics}, \text{Appeal}, \text{Exit} \rangle. \tag{3.1}$$

A mission without non-objectives is especially dangerous because an optimizer can satisfy a metric by changing what the institution was supposed to protect. A mission without appeal and exit may become domination even if its nominal objective is benevolent.

3.3 MONTREAL.AI: the institution layer

MONTREAL.AI is the builder, parent institution, and constitutional steward. It turns a mission into a governed object through GoalOS, Proof Gradient, AGI Jobs, SUCCESSOR Ω, MCAP specifications, and deployment controls. Its highest-value product is not a single model. It is customer-owned mission intelligence: a portable institution containing the customer's Mission Constitution, world programs, evidence bindings, tests, policies, admitted memory, Chronicle, and deployment controls.

The builder's role must be powerful but not unlimited. MONTREAL.AI can author the reference architecture and operate implementations while still being ineligible to self-certify every consequential claim. A credible institution can preserve founder stewardship of the parent while separating final verification, customer acceptance, and consequence-bearing admission according to risk.

3.4 AGI.eth: the action and economic layer

AGI.eth is the addressable root of agents and machine work. It can identify agents, validators, jobs, tools, nodes, missions, and settlement surfaces. The root is useful because it replaces opaque hexadecimal addresses and vendor-specific accounts with semantic identities that can resolve to controller disclosures, endpoints, versions, evidence, and status.

The economic layer is not merely payment. It is a mechanism for exposing commitments. A funded job records what value is at stake. An agent bond records willingness to bear a defined consequence. A validator bond makes evaluation strategically consequential. An escrow contract makes the settlement rule inspectable. A completion NFT can serve as a receipt, but only if the underlying job, evidence, and settlement remain discoverable.

The economic loop is:

$$\begin{aligned}\text{Mission} &\to \text{Job} \to \text{Escrow} \to \text{Execution} \to \text{Evidence} \\ &\to \text{Validation} \to \text{Acceptance/Dispute} \to \text{Settlement} \to \text{Chronicle}.\end{aligned} \tag{3.2}$$

Each arrow is a distinct transition. Execute, validate, accept, admit, and settle are not synonyms.

3.5 ASI.eth: the addressable civilization layer

ASI.eth is not claimed to be ASI. It is the proposed addressable root of a Machine Civilization architecture. The root should allow an independent client to discover:

- the canonical Root Manifest and protocol version;
- the Successor Covenant Ω;
- the namespace and role registry;
- the current or nominated successor and its status;
- proof requests, frozen candidates, and results;
- verifier identities, controllers, conflicts, and methods;
- admission decisions and authority envelopes;
- Chronicle events, failures, repairs, revocations, and succession;
- material limitations and non-claims.

The canonical namespace beneath ASI.eth can include successor.asi.eth, covenant.asi.eth, registry.asi.eth, chronicle.asi.eth, proof.asi.eth, verifier.asi.eth, admission.asi.eth, policy.asi.eth, rollback.asi.eth, and treasury.asi.eth. Names should be activated only when the corresponding functions are real and their controllers are disclosed.

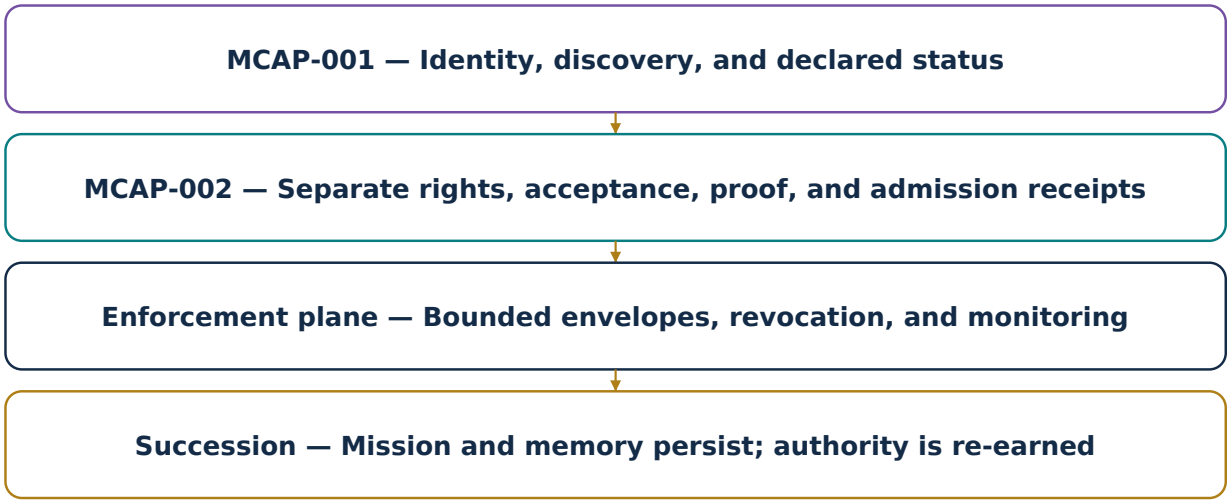


Figure 3.2: The protocol stack: addressability, economic work, evidentiary admission, authority, and succession are layered rather than conflated.

3.6 The institutional flywheel

The full architecture forms a positive feedback loop:

$$\begin{aligned}
\text{Mission legitimacy} &\to \text{qualified demand}, && (3.3)\\
\text{qualified demand} &\to \text{governed deployments}, && (3.4)\\
\text{deployments} &\to \text{jobs, evidence, settlement}, && (3.5)\\
\text{evidence} &\to \text{proof, admission, memory}, && (3.6)\\
\text{admitted memory} &\to \text{better successors}, && (3.7)\\
\text{better successors} &\to \text{greater mission value}. && (3.8)
\end{aligned}$$

The loop is beneficial only if the proof and authority gates remain binding. Without them, accumulated memory can amplify uncorrected error; economic incentives can reward collusion; and a strong brand can substitute for evidence. The flywheel therefore has a brake at every consequential transition.

Design Principle 3.1 (Semantic scarcity)**.** Use a small number of canonical identities with precise functions. Reserve additional names where strategically useful, but activate only those whose role, controller, status, and evidence are inspectable. Semantic scarcity increases institutional clarity.

3.7 Ownership without domination

Customer ownership is compatible with a common protocol. A customer should be able to export its mission, admitted memory, evidence references, policies, Chronicle, and successor state without acquiring MONTREAL.AI's generic software, marks, or root assets. Conversely, MONTREAL.AI can retain the architecture while lacking unilateral authority over a customer's consequential operations.

This separation resembles a constitutional platform: the common grammar is shared, while each mission institution has its own data, rights, evidence, acceptance, controllers, and authority. The result avoids both vendor lock-in and uncontrolled public ownership. Portability is a designed right; authority remains mission-specific.

4 Institutional State and the Successor Operator

PART II

FORMAL THEORY OF SUCCESSION

4.1 The typed institutional state

Let the complete state of a mission institution at time t be

$$X_t = \big(\mathcal{M}_t, \mathcal{K}_t, \mathcal{E}_t, \mathcal{P}_t, \mathcal{A}_t, \mathcal{C}_t, \mathcal{R}_t, \Theta_t\big). \tag{4.1}$$

The components are deliberately typed:

- $\mathcal{M}_t$: mission constitution, objectives, non-objectives, obligations, and metrics;
- $\mathcal{K}_t$: admitted memory and validated knowledge;
- $\mathcal{E}_t$: evidence graph, proof plans, receipts, and limitations;
- $\mathcal{P}_t$: rights, legal, security, policy, and constitutional constraints;
- $\mathcal{A}_t$: current authority envelope and permissions actually issued;
- $\mathcal{C}_t$: Chronicle state and commitments;
- $\mathcal{R}_t$: roles, controllers, beneficial-control disclosures, and revocation paths;
- Θ_t: the active implementation: models, prompts, tools, adapters, code, and runtime configuration.

The types matter because an institution fails when an object crosses a boundary without the required transformation. A model output cannot enter $\mathcal{K}$ merely by being copied into a database; it must be admitted through an evidence and rights process. A proof result cannot become $\mathcal{A}$ merely by exceeding a numerical threshold; a Principal must issue a bounded authority envelope. A transaction hash cannot become customer acceptance; an authorized customer must sign a separate receipt.

4.2 The succession operator

A successor candidate Θ_{t+1} is created from a predecessor state through a succession operator S. Define the set of inheritable components

$$\mathcal{I}_t = \big(\mathcal{M}_t, \mathcal{K}_t^{\mathrm{adm}}, \mathcal{O}_t, \mathrm{Refs}(\mathcal{C}_t)\big), \tag{4.2}$$

where $\mathcal{O}_t$ denotes surviving obligations and $\mathcal{K}_t^{\mathrm{adm}}$ is the subset of memory whose evidence, rights, relevance, and retention status remain valid. The pre-admission successor state is

$$\mathsf{S}(X_t, \Theta_{t+1}) = \big(\mathcal{M}_t, \mathcal{K}_t^{\mathrm{adm}}, \mathcal{E}_t^{\mathrm{hist}}, \mathcal{P}_{t+1}, \emptyset, \mathcal{C}_t \,\|\, c_{\mathrm{nom}}, \mathcal{R}_{t+1}, \Theta_{t+1}\big). \tag{4.3}$$

The symbol ‖ denotes Chronicle append. Historical evidence may be inherited as context but is not current proof of the new implementation. The authority component is reset to the empty envelope ∅.

Proposition 4.1 (Authority noninheritance)**.** *Assume every consequential action requires a valid authority envelope signed by an authorized Principal, every envelope binds the subject implementation digest and generation, and a new admission requires fresh proof of that frozen subject. If the succession operator sets $\mathscr{A}_{t+1} = \emptyset$ and changes the subject digest from d_t to $d_{t+1} \neq d_t$, then the successor cannot validly exercise predecessor authority without a new admission.*

Proof. Any predecessor envelope binds d_t. By assumption, authorization checks require equality between the executing subject digest and the digest in the envelope. The successor presents $d_{t+1} \neq d_t$, so the old envelope fails subject binding. Because $\mathscr{A}_{t+1} = \emptyset$, no current envelope exists. Therefore a new proof and Principal-signed admission are necessary before consequential authority can be issued. The result depends on correct signature verification, subject binding, and enforcement at the action gateway; it is not achieved by documentation alone. □

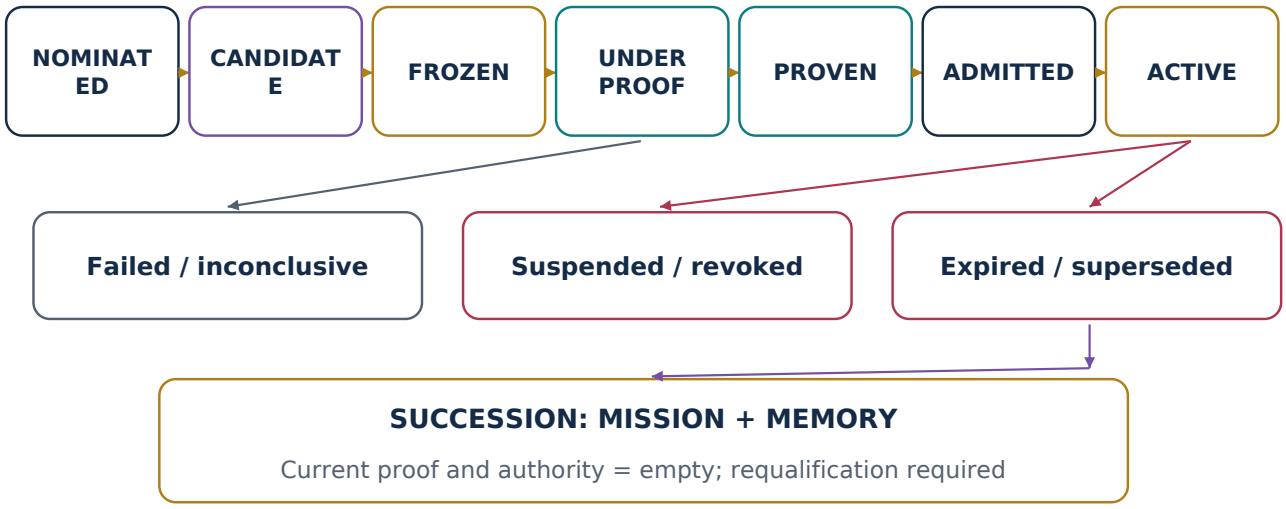


Figure 4.1: Successor state machine. The mission and admitted memory can cross the generational boundary; current authority cannot.

4.3 Memory admission

Raw memory m is admitted only if it satisfies a predicate

$$\mathsf{AdmitMemory}(m) = \mathsf{Integrity}(m) \wedge \mathsf{Rights}(m) \wedge \mathsf{Evidence}(m) \wedge \mathsf{Scope}(m) \wedge \mathsf{Retention}(m). \tag{4.4}$$

Each term can have degrees, but the conjunction is useful operationally: one failed hard gate prevents admission. The institution may retain rejected material in a quarantined evidence store for audit or repair, but it does not become trusted memory.

Let w_i be a memory item's proof weight, r_i its relevance to the current mission, q_i its data quality, and s_i a staleness penalty. A soft retention score can be defined as

$$\rho_i = w_i r_i q_i e^{-\lambda s_i}, \tag{4.5}$$

subject to the hard rights and integrity gates. High relevance cannot compensate for absent rights; high proof quality cannot compensate for a corrupted artifact.

4.4 Versioned identity and lineage

Persistent identity is represented by a root identifier n and a lineage graph

$$G_{\text{lin}} = (V_\Theta, E_{\text{succ}}), \tag{4.6}$$

where vertices are immutable implementation commitments and directed edges are succession events. A human-readable name may resolve to the current vertex, but the Chronicle preserves prior vertices. This prevents a mutable name from erasing history.

A lineage edge contains at least

$$\ell_{t\to t+1} = \langle d_t, d_{t+1}, h(\mathscr{I}_t), \mathsf{P}_{t+1}, \mathsf{Adm}_{t+1}, \mathscr{A}_{t+1}, c_{t+1} \rangle. \tag{4.7}$$

The edge is not an assertion that the two implementations are identical. It asserts a governed continuation under a declared inheritance and qualification process.

4.5 Institutional equivalence

Two implementations may be considered equivalent for a mission only relative to an observational and authority context. Let $\mathscr{T}$ be the declared test distribution, $\mathscr{U}$ the permitted action set, and ε a tolerance. Define

$$\Theta_i \sim_{\mathscr{M},\mathscr{T},\mathscr{U},\varepsilon} \Theta_j \tag{4.8}$$

when their mission-relevant outcome distributions are within ε, their policy violations are equivalent under the declared risk rule, and neither requires authority outside $\mathscr{U}$. This is a local, falsifiable equivalence, not a claim of general sameness.

The definition prevents a common succession error: a new model may score similarly on average while producing a different tail-risk distribution, requiring different tools, or behaving differently under adversarial prompts. Institutional equivalence therefore includes failure modes, not merely central performance.

4.6 Mission advantage

A successor should not be admitted merely because it is newer. Let $J(\Theta)$ denote expected mission value after costs and failures. Define the Mission Advantage of candidate Θ_c relative to the stronger of Current and Beta baselines:

$$\alpha(\Theta_c) = J(\Theta_c) - \max\{J(\Theta_{\text{current}}), J(\Theta_\beta)\} - C_{\text{switch}} - R_{\text{tail}} - F_{\text{frag}}. \tag{4.9}$$

Penalties must be incremental: do not count a cost or risk twice if it is already deducted in J. Here C_{switch} is transition cost, R_{tail} is residual tail risk, and F_{frag} is a fragility penalty capturing dependence on vendors, hidden state, unavailable data, or irreproducible tools. A positive α is necessary but not sufficient. Rights, proof, authority, and rollback gates still apply.

Design Principle 4.2 (No novelty premium)**.** Model recency, parameter count, brand prestige, and benchmark reputation are not proof of mission advantage. The candidate must outperform the better available baseline after cost, risk, fragility, and requalification burden.

5 Proof, Admission, and Authority

5.1 Evidence as a typed graph

Evidence is represented as a directed acyclic graph

$$G_{\mathscr{E}} = (V_{\mathscr{E}}, E_{\text{sup}}, E_{\text{conf}}, E_{\text{dep}}), \tag{5.1}$$

whose nodes include claims, artifacts, tests, observations, signatures, rights records, and limitations. Support edges indicate that an artifact bears on a claim; conflict edges record contradiction; dependency edges expose common sources that would otherwise be mistaken for independent corroboration.

This graph structure addresses a frequent failure in assurance systems: ten reports generated from one hidden source are not ten independent pieces of evidence. The effective evidence weight must discount dependence.

Let w_i be the quality weight of artifact i and Σ its dependency correlation matrix. One conservative effective-evidence measure is

$$W_{\text{eff}} = \frac{(1^\top \mathrm{w})^2}{\mathrm{w}^\top \Sigma \mathrm{w} + \epsilon}. \tag{5.2}$$

The formula is not a universal legal or statistical standard. It forces reviewers to represent correlated evidence rather than count documents.

5.2 Evidence classes

MCAP-002 defines five evidence classes:

Table 5.1: Evidence classes and promotion requirements.

Class	Name	Meaning	Promotion requirement
E0	Declared	Attributable assertion without corroborating evidence	Add source, version, and inspectable artifact
E1	Source-grounded	Evidence is attributable but not independently replayed	Independent reproduction from disclosed inputs
E2	Reproducible	A third party can reproduce the result	Independence of controller, incentives, and evidence custody
E3	Independently controlled	Verifier, customer, or custodian is outside claimant control and conflicts are disclosed	Measurable, rights-cleared consequence under bounded authority
E4	Consequence-bearing	Accepted outcome with observable consequence and governance	Repetition, monitoring, and successful succession

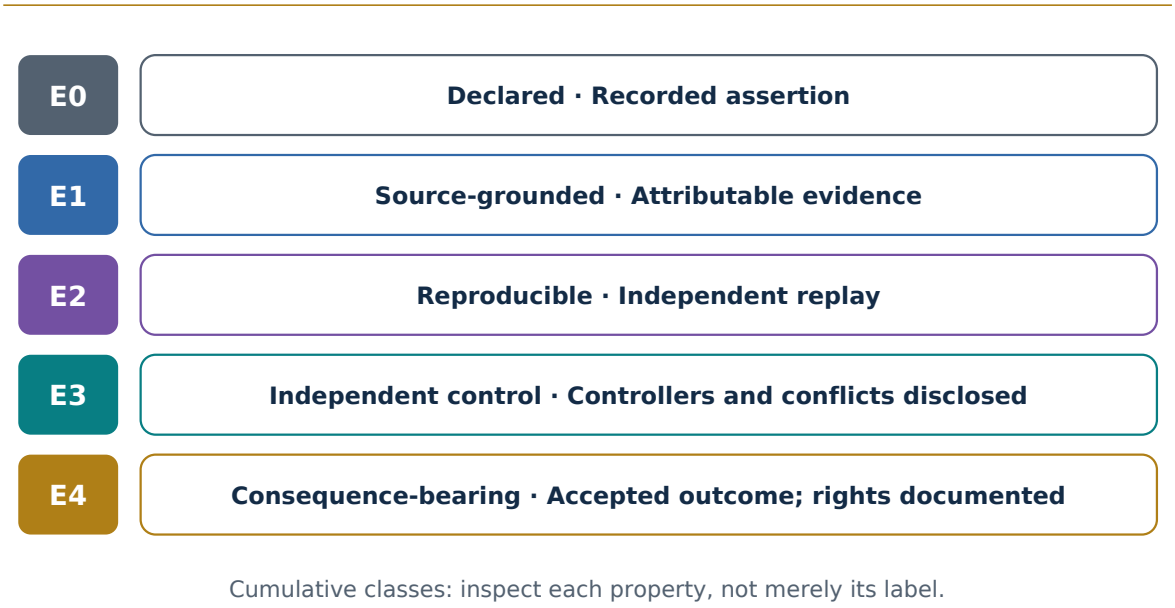


Figure 5.1: Evidence maturity is not a rhetorical scale. Each transition requires a new property that the prior class does not supply.

A hash, token, ENS name, NFT, or timestamp can strengthen integrity and attribution, but none automatically changes the evidence class. An E0 assertion hashed on-chain remains an E0 assertion with strong timestamping. A protocol-completed job may be E1 even when its transaction is indisputable, because off-chain rights, quality, and independence remain unproven.

5.3 The proof packet

A proof packet `P` should bind:

$$\mathsf{P} = \langle \mathscr{M}, d_\Theta, T, D, M, B, E, L, V, S, t_0, t_1 \rangle, \tag{5.3}$$

where d_Θ is the frozen candidate digest, T the test plan, D the complete denominator, M metrics, B baselines, E evidence, L limitations, V verifier identity and conflicts, S signatures, and $[t_0, t_1]$ the validity interval.

The denominator D is critical. If a system reports only the cases it chose to run or only the successes it retained, the posterior is biased. A complete-denominator proof records all committed cases, exclusions, failures, abstentions, and invalid runs.

Freeze-before-proof rule

The final proof plan and candidate digest must be committed before decisive evaluation begins. Post-hoc changes require either a new proof version or an explicitly separated exploratory analysis. The claimant may improve a future candidate; it may not silently improve the candidate under evaluation.

5.4 Bayesian interpretation

Let H_1 be the claim that a candidate satisfies the mission threshold and H_0 its relevant alternative. Evidence e updates log odds by

$$\log \frac{P(H_1 \mid e)}{P(H_0 \mid e)} = \log \frac{P(H_1)}{P(H_0)} + \log \frac{P(e \mid H_1)}{P(e \mid H_0)}. \tag{5.4}$$

Independent reproduction matters because correlated evidence contributes less new likelihood information. Precommitment matters because adaptive selection can inflate $P(e \mid H_0)$. Limitations matter because the posterior is conditional on the tested distribution; it does not automatically generalize to unobserved environments.

The value of proof can be defined decision-theoretically:

$$\mathrm{VoP}(\mathsf{P}) = \mathbb{E}\big[U(d^*(\mathsf{P}), \theta)\big] - \mathbb{E}\big[U(d_0, \theta)\big] - C(\mathsf{P}), \tag{5.5}$$

where $d^*(\mathsf{P})$ is the optimal decision after proof, d_0 the decision without it, and $C(\mathsf{P})$ the cost of evidence. The best proof is not necessarily the largest. It is the smallest evidence program that changes a consequential decision with acceptable residual uncertainty.

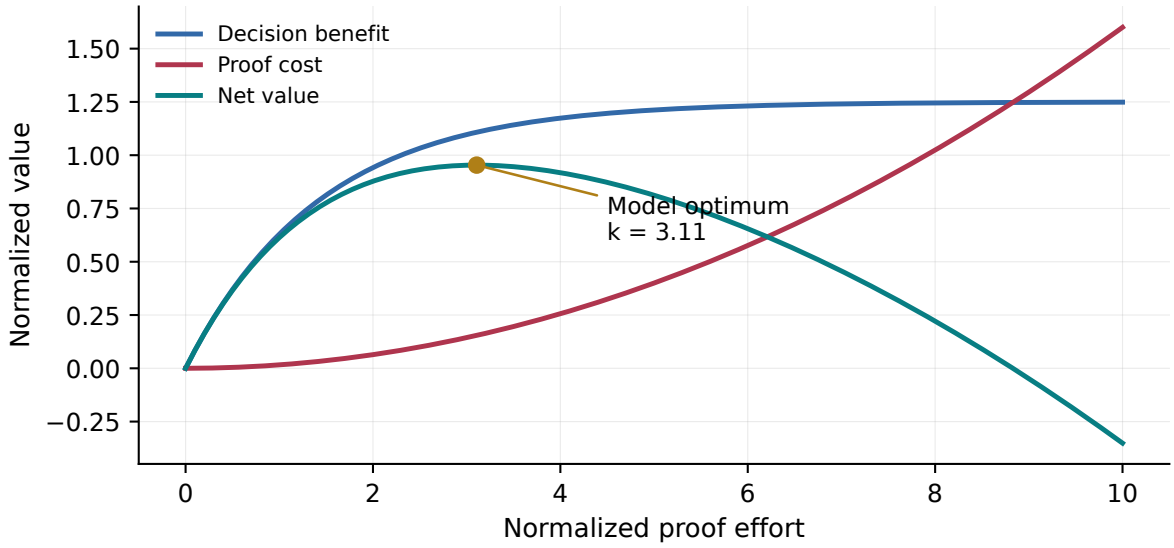


Figure 5.2: Specified proof-effort model: benefit $1.25(1-e^{-0.7k})$, cost $0.016k^2$, and their difference. The optimum is conditional on these assumed functions; it is not an estimate of customer value.

5.5 Admission as a distinct act

Proof states what the evidence supports. Admission states what the institution is willing to authorize. Let

$$\mathsf{Adm} : (\mathsf{P}, \mathscr{P}, \mathscr{M}, \mathscr{R}) \mapsto \{\text{reject}, \text{defer}, \text{admit}(\mathscr{A})\}. \tag{5.6}$$

Admission can be more conservative than proof. A candidate may demonstrate A3-level technical capability while receiving only A1 recommendation authority because value at risk, legal uncertainty, or rollback capacity is insufficient.

5.6 The authority envelope

An authority envelope is

$$\mathscr{A} = \langle s, m, U, T, B, \tau_0, \tau_1, O, R, K, \mathsf{P}, \mathsf{Adm}, \sigma \rangle, \tag{5.7}$$

where s is subject identity, m mission, U allowed actions, T targets, B quantitative bounds, $[\tau_0, \tau_1]$ validity, O observability requirements, R revocation, K rollback, proof and admission references, and Principal signature σ.

Define an authority partial order $\preceq$ where $\mathscr{A}_1 \preceq \mathscr{A}_2$ if every action, target, bound, and time interval in $\mathscr{A}_1$ is contained in $\mathscr{A}_2$. The institutional rule is

$$U(\mathscr{A}) \subseteq U(\mathscr{Q}(\mathsf{P})) \cap U(\mathscr{M}) \cap U(\mathscr{P}), \tag{5.8}$$

where $U(\cdot)$ maps each constraint to its allowed action-context set, $\mathscr{Q}(\mathsf{P})$ is the authority supported by proof, $\mathscr{M}$ the mission ceiling, and $\mathscr{P}$ the policy ceiling.

Proposition 5.1 (Authority monotonicity with hard ceilings). *Suppose* $\mathsf{P}_1 \preceq_E \mathsf{P}_2$ *means that* P_2 *weakly dominates* P_1 *in evidence quality, scope, recency, and independence. If* $\mathcal{Q}$ *is monotone, then* $\mathcal{Q}(\mathsf{P}_1) \preceq \mathcal{Q}(\mathsf{P}_2)$. *Nevertheless, admitted authority cannot exceed the fixed mission and policy ceilings in equation (5.8).*

Proof. Monotonicity gives the first relation. The admitted envelope must be a subset of the intersection of evidence-supported, mission-permitted, and policy-permitted action-context sets. It need not be the maximal such set. Enlarging one operand cannot enlarge the meet beyond either of the remaining operands. More proof can justify more authority only where mission and policy already allow it.

For a request a, these are sets of permitted action-context pairs, not scalar scores:

$$U_{\text{adm}} \subseteq U_{\text{proof}} \cap U_{\text{mission}} \cap U_{\text{rights}} \cap U_{\text{policy}} \cap U_{\text{risk}}. \tag{5.9}$$

A gateway must evaluate membership using one complete valid envelope. It must not combine the target of one grant with the action or budget of another. □

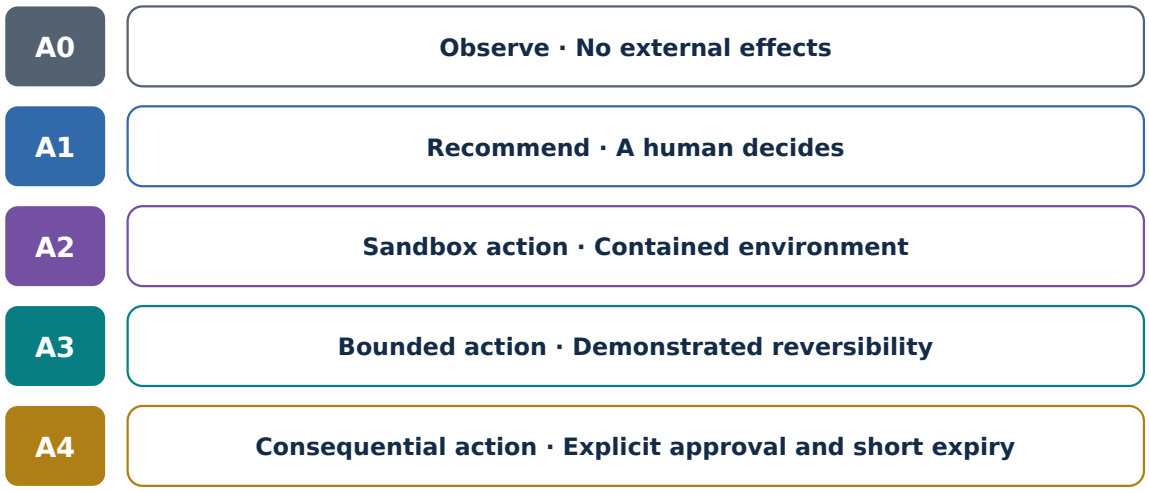


Figure 5.3: Default authority ladder. Movement upward requires stronger evidence and stronger control; succession returns the successor to A0 until requalification.

5.7 Default authority ladder

The default ladder is:

- A0 Observe: inspect, resolve, and simulate without issuing operational recommendations or external actions.
- A1 Recommend: issue attributable recommendations to an accountable human or institution.
- A2 Sandbox action: act in a bounded test environment with no uncontrolled external effect.
- A3 Reversible bounded action: perform production actions within quantitative and target limits, with monitoring and tested rollback.
- A4 Explicit consequential action: execute a specifically approved high-consequence action with short expiry and direct accountable authorization.

The v9 reference package supplies A0 formation and A1/A2 rehearsal. External consequential effects remain disabled. This is an architectural boundary, not an omission to be hidden.

6 Chronicle, Viability, and Rollback

6.1 The Chronicle as institutional memory of consequence

A database records state. A Chronicle records why consequential state was allowed to change. It is an append-only sequence of commitments to admissions, actions, failures, repairs, suspensions, revocations, and successions. Sensitive evidence may remain protected, but its existence, custodian, digest, status, and use in a decision should be attributable.

Let event c_i contain a canonical message m_i, timestamp t_i, signer set S_i, evidence references E_i, and authority state A_i. Define the chain

$$h_i = H(\text{domain} \parallel h_{i-1} \parallel \text{canon}(m_i, t_i, S_i, E_i, A_i)), \tag{6.1}$$

where H is a collision-resistant hash function and `domain` prevents cross-protocol ambiguity. Events can additionally be batched into a Merkle root and independently timestamped [24, 40].

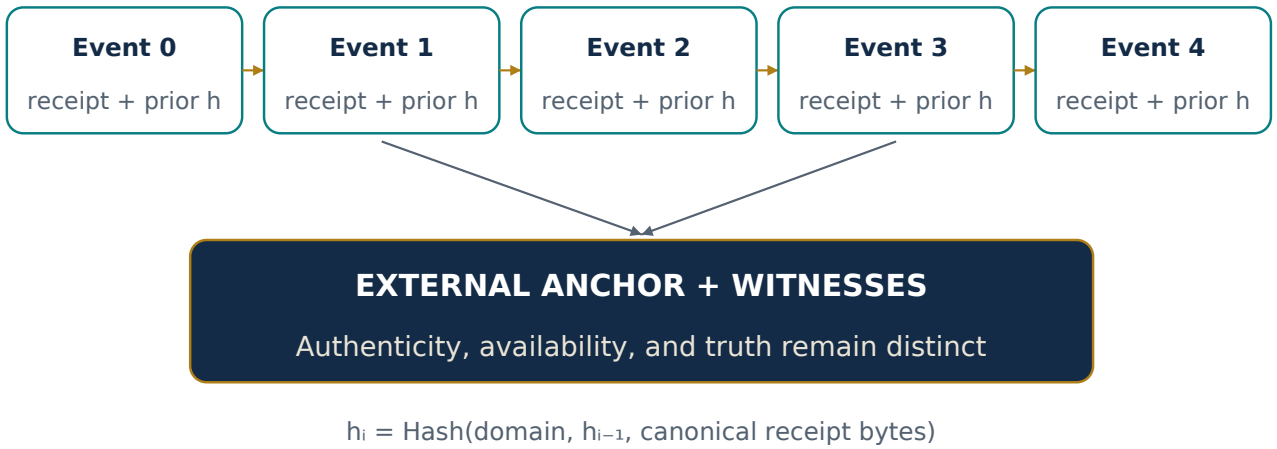


Figure 6.1: A Chronicle links institutional transitions, not merely files. External anchoring and separate custody strengthen resistance to retrospective rewriting.

Proposition 6.1 (Tamper evidence)**.** *Assume canonical serialization is deterministic and H is collision resistant. Altering any accepted event c_j without reproducing all later commitments changes h_j and therefore every subsequent digest. If at least one later digest is retained by an independent custodian, undetected retrospective alteration is computationally infeasible under the hash assumption.*

Proof. The changed canonical message changes the input to H at index j, except with a hash collision. This changes h_j, which is an input to h_{j+1}, and induction changes all later digests. An independent copy of a later digest therefore detects divergence. The proposition proves tamper evidence, not truth of the original message. □

6.2 Chronicle event taxonomy

A useful event taxonomy includes:

- `MISSION_CONSTITUTED`;
- `SETTLEMENT_IMPORTED`;
- `RIGHTS_REVIEWED`;
- `CUSTOMER_ACCEPTED`;
- `CANDIDATE_FROZEN`;
- `PROOF_STARTED`, `PROOF_PASSED`, `PROOF_FAILED`, `PROOF_INCONCLUSIVE`;
- `ADMISSION_GRANTED`, `ADMISSION_REJECTED`;
- `AUTHORITY_ISSUED`, `AUTHORITY_SUSPENDED`, `AUTHORITY_REVOKED`, `AUTHORITY_EXPIRED`;
- `ROLLBACK_INITIATED`, `ROLLBACK_VERIFIED`;
- `MEMORY_ADMITTED`, `MEMORY_QUARANTINED`;
- `SUCCESSOR_NOMINATED`, `SUCCESSOR_QUALIFIED`, `SUCCESSOR_SUPERSEDED`.

The taxonomy preserves distinct causality. An authority event references an admission; an admission references a proof; a proof references a frozen candidate and complete denominator; a successor qualification references fresh proof rather than predecessor reputation.

6.3 Viability under bounded authority

Consider controlled institutional dynamics

$$x_{t+1} = f(x_t, u_t, w_t), \tag{6.2}$$

where x_t includes operational and governance state, u_t is an authorized action, and $w_t \in W$ is a disturbance. Let the policy-safe set be

$$K = \{x \,:\, g_j(x) \le 0,\ j = 1, \dots, m\}. \tag{6.3}$$

The **viability kernel** is the set of states from which there exists an admissible policy keeping the system in K for all modeled disturbances [5]. The authority envelope should restrict actions to an inner approximation $\widehat{K} \subseteq K$ that can be verified in real time.

A barrier function $B(x)$ can enforce safety when

$$B(x) \le 0 \Rightarrow \exists u \in U_{\mathscr{A}}(x) \text{ such that } B(f(x,u,w)) \le 0 \quad \forall w \in W. \tag{6.4}$$

The practical interpretation is direct: before an action is executed, the gateway evaluates whether the next state remains within the authorized safe set under modeled uncertainty. When this cannot be established, the system abstains or requests a narrower action.

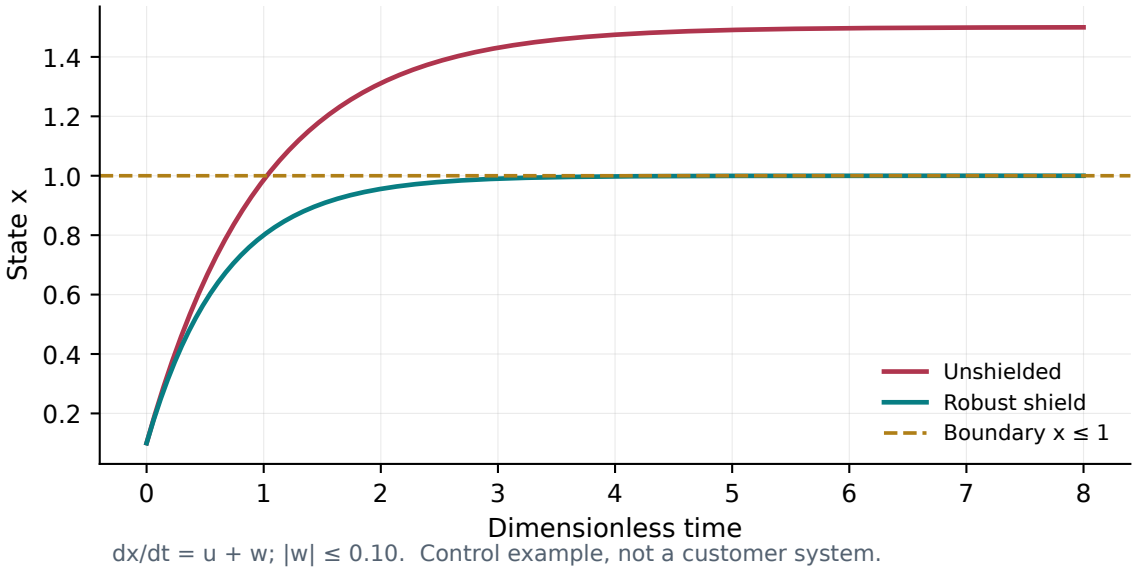


Figure 6.2: A specified robust-control example with $\dot{x} = u + w$, $|w| \leq 0.10$, and safe set $x \leq 1$. The shield enforces $u \leq 1.5(1 - x) - 0.10$ when feasible. This trajectory is not a computed viability kernel for a customer system.

6.4 Lyapunov-style recovery arguments

Let x^* be a safe operating equilibrium. A Lyapunov candidate $V(x)$ supports a local stability argument if

$$V(x) > 0 \quad (x \neq x^*), \qquad \Delta V(x) = V(f(x, \pi(x), w)) - V(x) < 0 \tag{6.5}$$

for the nominal equilibrium (or disturbances that vanish there), within a declared region. With persistent disturbances, practical or input-to-state stability requires a different bound; strict descent to zero is not automatic. Such bounds are local to the modeled system [31, 38]. In institutional deployments, V can aggregate policy deviation, unresolved exceptions, stale evidence, unreviewed actions, and distance from a recoverable state. The argument is not a substitute for tests; it specifies what the monitoring and controller must demonstrate.

A rollback path is adequate only if it has been exercised. Documentation stating that a prior version exists is not evidence that dependencies, credentials, databases, and external effects can be restored. A rollback receipt should record trigger, authority state, steps, elapsed time, residual effects, evidence loss, and corrective actions.

6.5 Irreversibility budget

Some effects cannot be rolled back: a public disclosure, physical shipment, irreversible transfer, or human decision may persist. Let $I(u, x)$ be the irreversibility cost of action u. The authority envelope should enforce

$$I(u, x) \leq I_{\max}(\mathscr{A}), \tag{6.6}$$

where $I_{\max}$ rises only with stronger proof and explicit accountable approval. A1 recommendations can have low technical irreversibility but high social influence; the model should include communication effects rather than treat "recommendation only" as consequence free.

Rollback is not absolution

Rollback can restore technical state while leaving economic, legal, reputational, or human consequences. A complete receipt distinguishes restored state from residual consequence. An institution should prefer actions whose downside can be bounded before execution, not merely actions it hopes to reverse later.

7 Mechanism Design and Strategic Stability

7.1 The institution as a mechanism

A Machine Civilization is a repeated mechanism among principals, customers, agents, verifiers, operators, custodians, and resource providers. Each participant has private information, distinct incentives, and possible conflicts. Governance cannot assume benevolence; it must make truthful, useful behavior strategically preferable where feasible.

Let agents $i \in N$ have types θ_i, reports r_i, allocation rule $g(r)$, transfer rule $p(r)$, and utility

$$u_i(\theta_i, r) = v_i\big(g(r), \theta_i\big) - p_i(r) - \ell_i(r), \tag{7.1}$$

where ℓ_i includes expected slashing, reputation loss, opportunity cost, and legal or contractual consequence. Mechanism design asks for rules under which desired outcomes arise despite strategic behavior [28, 50, 64].

7.2 Institutional desiderata

The following properties are mission-dependent but generally desirable:

Individual rationality. Participation under disclosed rules provides nonnegative expected utility relative to exit.

Incentive compatibility. Truthful reporting and competent execution dominate profitable manipulation under modeled conditions.

Solvency. Escrow, bonds, rewards, refunds, and treasury obligations remain fully accounted for.

No unilateral self-certification. A claimant cannot create the work, alter decisive evidence, serve as sole final verifier, grant itself authority, and erase the record.

Coalition resistance. Small coalitions cannot cheaply manufacture proof or capture admission.

Appeal and exit. Error correction does not depend on obedience to the actor that made the decision.

Bounded authority. Transfers and reputation may affect access but do not purchase unlimited permission.

No mechanism satisfies every desirable property under every information structure. The point is not to declare perfection but to publish the trade-offs and identify which failure modes remain profitable.

7.3 Separation-of-powers game

Consider a claimant C, verifier V, and Principal P. The claimant receives benefit b from admission, pays work cost c, and can spend manipulation cost m. The verifier receives reward r_v, incurs diligence cost d, and may receive bribe β. The Principal bears expected loss L from false admission.

In a captured design, C controls V or the evidence store, so manipulation succeeds with probability q_c. In a separated design, independent evidence custody, conflict disclosure, random assignment, and challenge reduce success probability to $q_s < q_c$ and increase expected detection penalty s.

Manipulation is unattractive when

$$q_s b - m - (1 - q_s)s < b - c_{\text{honest}}, \tag{7.2}$$

and verifier collusion is unattractive when

$$\beta - d_{\text{collude}} - p_{\text{detect}} s_v < r_v - d. \tag{7.3}$$

These inequalities show why "different wallets" are insufficient. Independence changes q_s, detection probability, and penalty only if controllers, incentives, evidence access, and challenge are genuinely separated.

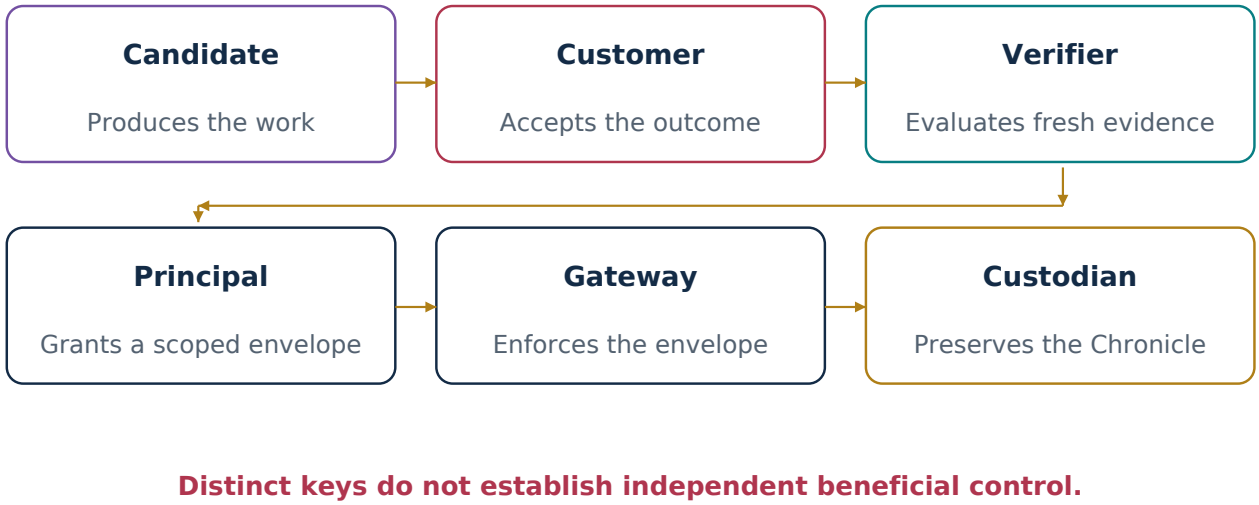


Figure 7.1: Role separation changes strategic payoffs only when control, evidence, incentives, and challenge are genuinely distinct.

7.4 Correlated equilibrium for distributed verification

A distributed verifier network need not converge through a single command node. Let $\pi(a)$ be a correlation device over verifier actions. It is a correlated equilibrium when, for every verifier i and deviation a_i',

$$\sum_{a_{-i}} \pi(a_i, a_{-i}) \left[u_i(a_i, a_{-i}) - u_i(a_i', a_{-i}) \right] \geq 0. \tag{7.4}$$

A public proof plan can serve as a correlation device: it tells participants what evidence to evaluate and how outcomes map to decisions, without requiring any verifier to know every other verifier's private signal. Correlated equilibrium does not guarantee truth. It supplies a tractable criterion for strategic consistency under recommendations [6].

Mechanism design should therefore separate three questions:

1. Are verifier strategies in equilibrium under the reward and penalty rules?

2. Does the equilibrium aggregate information accurately?
3. Is the resulting evidence sufficient for the proposed authority?

A mechanism can be stable and wrong. Collusive equilibria are possible. Independent ground truth, rotating verifier sets, adversarial challenge, and evidence reproducibility remain necessary.

7.5 Repeated cooperation

In a repeated Prisoner's Dilemma with reward R, temptation T, punishment P, and discount factor δ, a grim-trigger cooperative equilibrium is sustainable when

$$\frac{R}{1-\delta} \geq T + \frac{\delta P}{1-\delta} \quad \Longleftrightarrow \quad \delta \geq \frac{T-R}{T-P}. \tag{7.5}$$

The institutional design can increase effective δ by making identity persistent, preserving reputation evidence, and offering future opportunity. It can reduce T by limiting unilateral access and increase the cost of defection through bonds, loss of high-trust infrastructure, or contractual consequence. The objective is not punitive severity but credible repeated interaction [8].

7.6 Evolutionary dynamics

For strategy shares x_i and payoff matrix A, replicator dynamics are

$$\dot{x}_i = x_i \left[(Ax)_i - x^\top A x \right]. \tag{7.6}$$

A cooperative strategy grows when its payoff exceeds population average [39, 62]. Verification, reputation, escrow, and access control modify A; they do not guarantee cooperation by themselves. Poorly designed token rewards can make manipulation the high-payoff strategy.

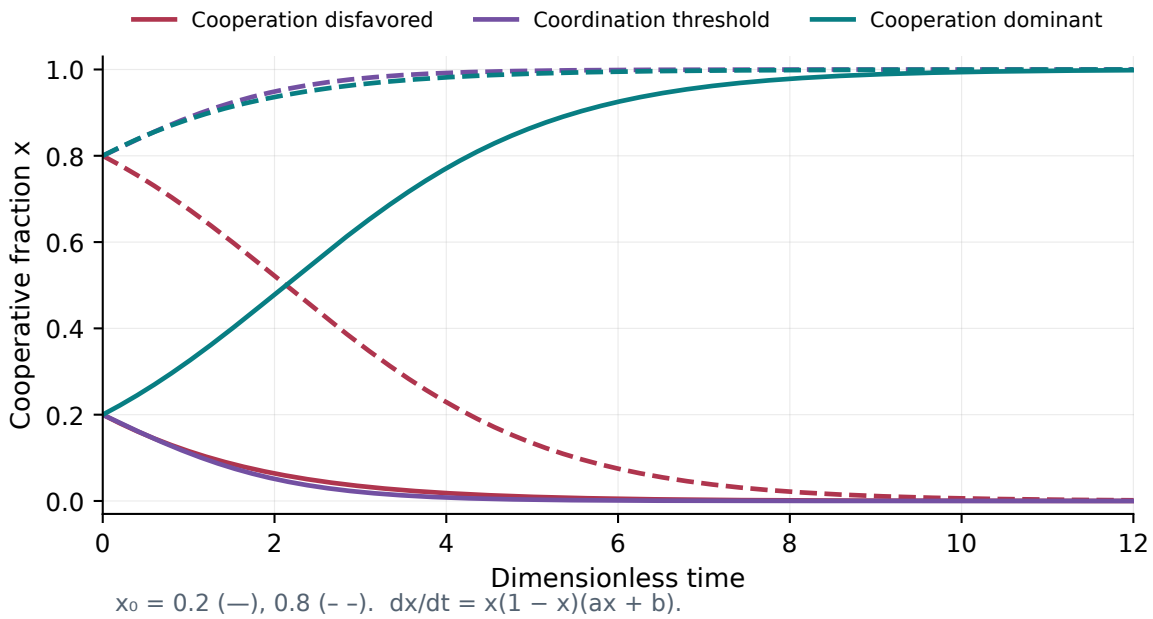


Figure 7.2: Specified two-strategy model: $\dot{x} = x(1-x)(ax+b)$. Solid and dashed curves begin at 0.2 and 0.8. The regimes $(a, b) = (0, -0.65), (2, -1), (0, 0.65)$ illustrate decline, a coordination threshold, and dominance. These are synthetic trajectories, not measured adoption.

Design Principle 7.1 (Make predation strategically shallow)**.** A robust institution does not assume predatory strategies disappear. It makes them observable, expensive, low-access, and unable to acquire persistent authority. Cooperative strategies should gain durable access to knowledge, reputation, insurance, and future missions only through evidence-bearing contribution.

7.7 Potential games and institutional repair

If a multi-agent interaction admits a potential function $\Phi(a)$ such that unilateral payoff changes align with changes in Φ, local learning can converge toward stable configurations [42]. A useful institutional potential can combine mission value, evidence quality, solvency, and recoverability:

$$\Phi = V_{\mathscr{M}} - \lambda_1 R_{\text{policy}} - \lambda_2 R_{\text{capture}} - \lambda_3 I + \lambda_4 Q_{\text{evidence}} + \lambda_5 B_{\text{rollback}}. \tag{7.7}$$

The weights are mission-specific and should not be hidden. A potential is useful when local incentives actually track it; otherwise it is merely a dashboard score. Repair consists of changing rules, access, or transfers so that locally profitable behavior again improves the institutional objective.

8 Statistical Physics of Institutional Stability

8.1 Three levels of claim

Statistical physics offers tools for distributions, fluctuations, collective order, and metastability. It does not establish that a social institution is a thermodynamic material. This chapter distinguishes an exact finite-state identity, a hypothetical stochastic process, and an institutional interpretation. Only the first is unconditional within its mathematical definitions. Behavioral predictions require a transition model and empirical calibration; the institutional interpretation is a research hypothesis.

8.2 Admissible institutional ensembles

Let s encode a finite configuration of agents, evidence, roles, obligations, and permissions. Define a cost function in one declared unit, for example normalized mission-loss units:

$$H(s) = C(s) + \lambda_R R(s) + \lambda_I I(s) + \lambda_C C_{\text{capture}}(s) - \lambda_V V_{\mathscr{M}}(s) - \lambda_Q Q_{\mathscr{E}}(s). \tag{8.1}$$

The coefficients must convert heterogeneous quantities to that unit. A low value is a modeled preference, not moral legitimacy. Constitutional prohibitions define a nonempty support $\mathcal{S}_{\text{adm}} \subseteq \mathcal{S}$ independently of this score. Assigning a large finite penalty does not enforce a prohibition: a positive-temperature Gibbs distribution gives every finite-cost supported state positive probability.

For $\tau > 0$ in the same unit as H, put $\beta = \tau^{-1}$ and define

$$Z_\tau = \sum_{s \in \mathcal{S}_{\text{adm}}} e^{-H(s)/\tau}, \qquad p_\tau(s) = Z_\tau^{-1} e^{-H(s)/\tau}, \qquad F_\tau = -\tau \log Z_\tau. \tag{8.2}$$

The exponent is dimensionless. Here τ is an effective selection-noise parameter, not a measured temperature in kelvin. The canonical-ensemble analogy is inspired by statistical mechanics [21].

8.3 Finite-state Gibbs variational principle

For a probability distribution q supported on $\mathcal{S}_{\text{adm}}$, define

$$\mathscr{F}_\tau[q] = \sum_s q(s) H(s) + \tau \sum_s q(s) \log q(s), \tag{8.3}$$

with $0 \log 0 = 0$.

Proposition 8.1 (Finite-state variational identity)**.** *If $\mathcal{S}_{\text{adm}}$ is finite and nonempty, H is finite on it, and $\tau > 0$, then*

$$\mathscr{F}_\tau[q] - F_\tau = \tau D_{\text{KL}}(q \| p_\tau) \geq 0. \tag{8.4}$$

The unique minimizer is $q = p_\tau$.

Proof. Substitute $\log p_\tau(s) = -H(s)/\tau - \log Z_\tau$ into $\tau \sum_s q(s) \log[q(s)/p_\tau(s)]$ and use $\sum_s q(s) = 1$. Nonnegativity of relative entropy gives the inequality; equality requires equality of the distributions. The identity is algebraic. It does not prove that actual institutional behavior samples p_τ. □

8.4 A transition law is still necessary

On a connected finite admissible-state graph, let the proposal kernel $Q(s, s')$ be symmetric. A Metropolis acceptance probability

$$A(s, s') = \min\{1, e^{-[H(s')-H(s)]/\tau}\} \tag{8.5}$$

satisfies detailed balance with p_τ. Irreducibility plus an aperiodic self-loop gives convergence to that stationary law. Without these conditions, disconnected classes, deterministic cycles, or directed incentives can prevent the conclusion. Directed strategic institutions generally require nonequilibrium models rather than an assumed equilibrium law.

An exact potential game permits a related construction when asynchronous single-player logit updates use the same potential and noise scale. A designer cannot infer this property from a dashboard score; unilateral utility differences must actually match potential differences [42].

8.5 Entropy, diversity, and uncertainty

Use dimensionless Shannon entropy for institutional random variables [59]. If R represents control domains, E evidence state, and A action state, the chain rule is

$$\mathsf{H}(R, E, A) = \mathsf{H}(R) + \mathsf{H}(E \mid R) + \mathsf{H}(A \mid E, R). \tag{8.6}$$

This is not an additive decomposition into independently measurable "diversity," "uncertainty," and "opacity" scores. Joint dependence matters. A diverse verifier population may be desirable; uncertainty about its beneficial control may not be. Reducing total entropy is therefore not an adequate institutional objective.

8.6 Landau landscapes and metastability

For a signed order parameter m, consider the dimensionless quartic potential

$$F(m; a, h) = \frac{1}{4}m^4 + \frac{1}{2}am^2 - hm. \tag{8.7}$$

Here m is not itself a probability; a controls the landscape and h biases one direction. For $a = -1$ and $h = 0$, stationary points satisfy $m(m^2 - 1) = 0$: $m = \pm 1$ are minima and $m = 0$ is a maximum. The barrier height is $1/4$. This specifies a toy bistable system, not a measurement of cooperation in the product.

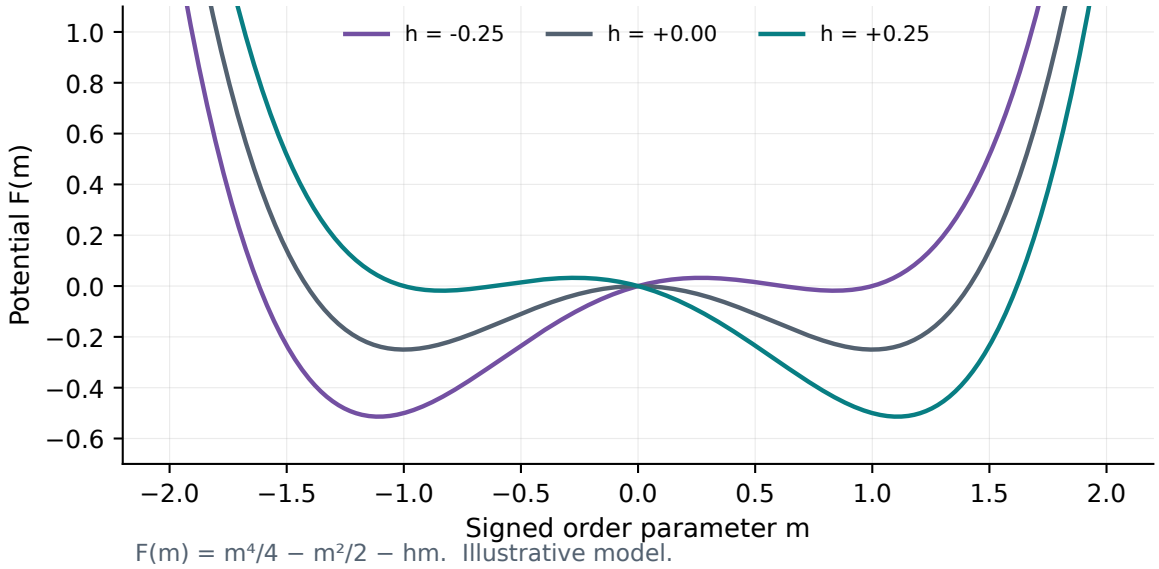


Figure 8.1: Specified quartic potentials $F(m) = m^4/4 - m^2/2 - hm$ for three bias fields. A basin can represent either beneficial coordination or capture, depending on how the state and objective are defined. Stability is not legitimacy.

8.7 Escape and controlled repair

For the one-dimensional gradient diffusion

$$dm_t = -U'(m_t)\,dt + \sqrt{2D}\,dW_t, \tag{8.8}$$

a small-noise escape approximation between nondegenerate wells has exponential dependence

$$k \propto e^{-\Delta U/D}. \tag{8.9}$$

The prefactor, boundary conditions, noise convention, and small-noise limit matter [20, 33]. This does not assign a numeric attack rate to a machine institution. The design analogy is to raise the cost of unauthorized capture while keeping authorized appeal, repair, and exit feasible. A high barrier that traps everyone in a bad equilibrium is not a safety achievement.

8.8 Finite systems and phase diagrams

For finitely many states with finite smooth costs and $\tau > 0$, Z_τ is a finite sum of positive terms. No thermodynamic singularity follows. Sharp crossovers, metastable switching, bifurcations in an explicit deterministic dynamics, and discontinuities in a zero-noise argmin are distinct phenomena.

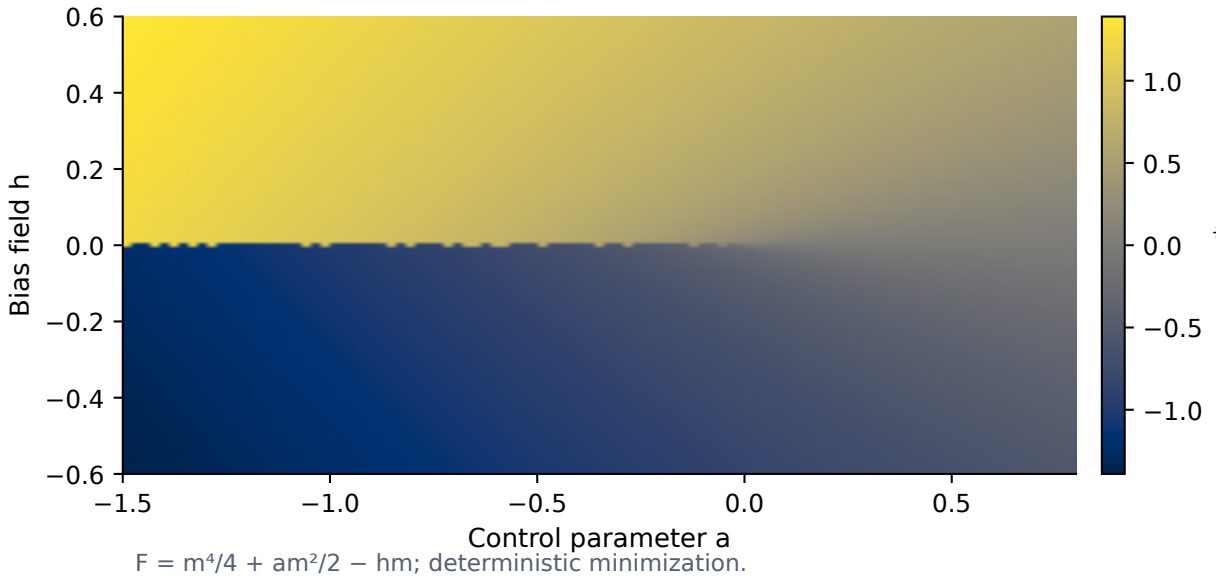


Figure 8.2: Deterministic minimizer of $m^4/4 + am^2/2 - hm$ on a declared grid. The plot is a minimizer map, not an empirical institutional phase diagram. At $h = 0, a < 0$, the two minima tie; grid tie-breaking selects one representative.

An institutional phase-transition claim would require a defined order parameter, an interaction graph, observations across scale or parameters, and tests against alternative explanations. Counting wallets does not establish independent sample size. For an exchangeable set with equal variance and pairwise correlation ρ, the familiar variance-equivalent size is $N_{\text{eff}} = N/[1 + (N-1)\rho]$ when the covariance model is valid. Common control can eliminate most of the apparent replication.

8.9 Gibbs versus Helmholtz: preserve the units

Physical Gibbs free energy is $G = H_{\text{enthalpy}} - TS$; thus

$$\Delta G = \Delta H_{\text{enthalpy}} - T\Delta S \tag{8.10}$$

holds between states at the same temperature. At variable temperature, one must retain $\Delta(TS)$. The physical Helmholtz potential is $F = U - TS$, and the canonical relation $F = -k_B T \log Z$ uses an appropriate physical ensemble. Equation (8.3) is an entropy-regularized finite-state cost functional; it is not measured institutional Gibbs energy.

What this model can and cannot establish

The variational identity is proved under explicit definitions. The numerical diagrams reproduce declared toy functions. An institutional interpretation requires an observable state mapping, units, estimated transition law, uncertainty, and falsification. Neither low modeled free energy nor a stable equilibrium confers rights, truth, or authority.

9 Hamiltonian Continuity, Information, and Causal Accountability

9.1 Continuity as constrained flow

Hamiltonian mechanics describes structured evolution in phase space [26]. For institutional modeling, let q represent configuration variables such as mission state, role assignments, memory structure, and deployment topology. Let p represent committed momentum: contracts, obligations, trained capabilities, accrued switching costs, and dependencies. A baseline Hamiltonian is

$$H(q, p, t) = T(p) + V(q, t), \tag{9.1}$$

with canonical flow

$$\dot{q} = \nabla_p H, \qquad \dot{p} = -\nabla_q H. \tag{9.2}$$

For a differentiable canonical solution,

$$\frac{dH}{dt} = \nabla_q H \cdot \nabla_p H - \nabla_p H \cdot \nabla_q H + \partial_t H = \partial_t H. \tag{9.3}$$

Thus H is conserved only in the autonomous case (or where $\partial_t H = 0$ along the solution). Canonical coordinates and conjugate variables require a justified state mapping; obligations are not automatically mechanical momenta. The analogy does not assert conservation of physical energy. It emphasizes that succession should preserve declared invariants while allowing implementations to change. Unmodeled commitments act like hidden momentum: they can drive the successor outside its intended mission even when its visible configuration appears correct.

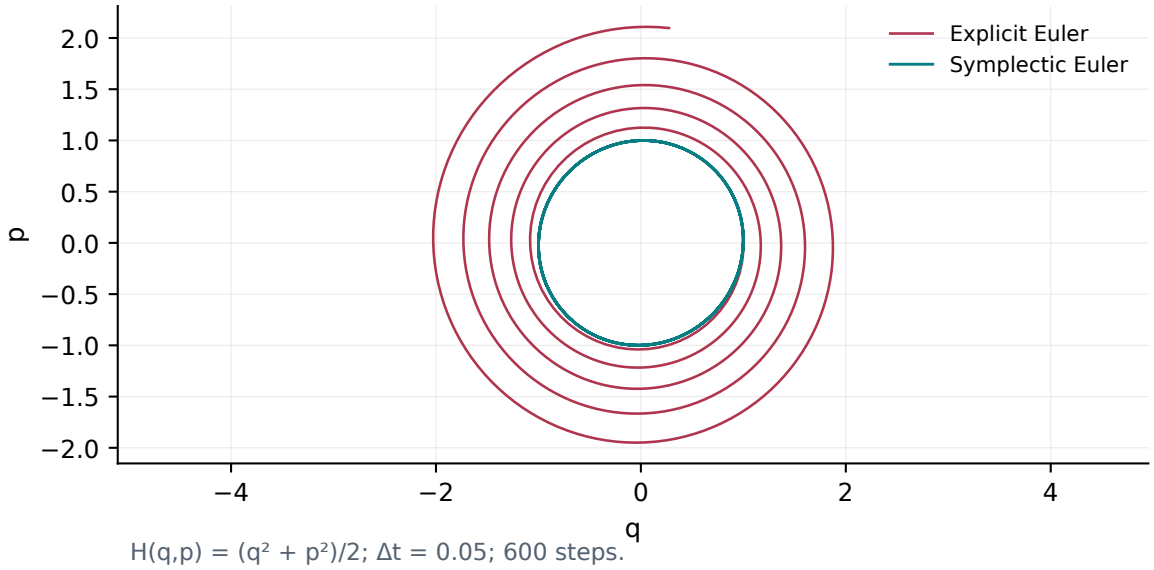


Figure 9.1: Specified harmonic oscillator $H = (q^2 + p^2)/2$: explicit and symplectic Euler trajectories at step 0.05 over 600 steps. Symplectic Euler has bounded oscillatory energy error here, not exact energy conservation. The comparison does not prove that a Chronicle is symplectic.

9.2 Port-Hamiltonian institutional dynamics

Real institutions are dissipative, controlled, and open. A port-Hamiltonian form is more appropriate:

$$\dot{z} = \big[J(z) - R(z)\big]\nabla H(z) + G(z)u + G_w(z)w, \tag{9.4}$$

where $J = -J^\top$ describes interconnection, $R = R^\top \succeq 0$ dissipation, u inputs, and w disturbances. For autonomous H, define $y = G^\top \nabla H$ and $y_w = G_w^\top \nabla H$. Then

$$\dot{H} = -\nabla H^\top R \nabla H + y^\top u + y_w^\top w. \tag{9.5}$$

This storage identity alone does not imply asymptotic stability. Positive definiteness around an equilibrium, output detectability, suitable feedback, and disturbance assumptions remain necessary [43]. Discrete expiry, revocation, and succession are hybrid jumps; they are not automatically terms in a continuous dissipation matrix. Institutional interpretations may associate dissipation with budget use or staleness only after defining variables and units. Ports include customer requests, proof results, policy changes, and economic settlement. The model makes a key point: external inputs can increase institutional energy or risk, so their interfaces require authorization and observability.

9.3 Constrained manifolds and constitutional invariants

Let constitutional constraints be $\phi_j(z) = 0$ and policy inequalities $g_k(z) \le 0$. The admissible set (not necessarily a smooth manifold) is

$$\mathcal{N} = \{z \,:\, \phi(z) = 0,\ g(z) \le 0\}. \tag{9.6}$$

An illustrative candidate update is

$$z_{t+1} = \Pi_{\mathcal{N}}(z_t + \Delta t\, f(z_t, u_t)), \tag{9.7}$$

where projection is single-valued only under additional conditions, such as a nonempty closed convex set in Euclidean geometry. A mixed discrete or nonconvex constitution requires a justified solver and post-update guard checks. Projection alone does not prove continuous-time invariance. Projection failure is a no-admit outcome; succession additionally resets signed authority through a separate discrete transition.

Typical invariants include:

- the current successor cannot erase the Chronicle alone;
- authority expires and binds a specific subject digest;
- customer acceptance and independent proof are separate signatures;
- rights failure blocks memory admission;
- successor authority starts empty;
- material implementation change triggers requalification.

9.4 Symplecticity and auditability

A symplectic integrator preserves the symplectic two-form for the numerical Hamiltonian map. It need not exactly conserve the original Hamiltonian. The institutional analogue is event-sourced transition design: rather than overwrite state, record the transition and derive current state from the accepted event sequence. The objective is not numerical symplecticity but reduced drift between declared rules and operational history.

A transition function

$$X_{t+1} = \mathsf{T}(X_t, e_t) \tag{9.8}$$

should be deterministic for a canonical event e_t, or explicitly record external nondeterminism. Replaying the Chronicle from a trusted checkpoint should reconstruct authority, admission, and successor status. If it cannot, the public state is not auditable.

9.5 Information-theoretic proof quality

Evidence is useful when it reduces uncertainty about a decision-relevant claim. Let Y be candidate competence under the mission and E observed evidence. Mutual information is

$$I(Y;E) = H(Y) - H(Y \mid E). \tag{9.9}$$

A large evidence packet can have low $I(Y;E)$ if it repeats correlated observations. Conversely, a carefully selected adversarial test can have high information gain. Proof-plan design can maximize expected information per cost:

$$T^* = \arg\max_T \frac{\mathbb{E}[I(Y;E_T)]}{C(T) + \lambda R(T)}. \tag{9.10}$$

The KL divergence

$$D_{\mathrm{KL}}(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \tag{9.11}$$

can measure distribution shift between proof and deployment environments. Authority should contract when shift exceeds a predeclared threshold, because proof relevance has decayed.

9.6 Causal accountability

Correlation between a successor and a good outcome is insufficient when many actors and systems intervene. A causal model represents variables and interventions through a directed acyclic graph [55]. Let Y be mission outcome, A the agent action, Z context, and U unobserved confounding. The causal effect

$$\mathbb{E}[Y \mid \mathrm{do}(A = a)] - \mathbb{E}[Y \mid \mathrm{do}(A = a')] \tag{9.12}$$

is identifiable only under declared assumptions or experimental design.

The Chronicle should therefore record not only the outcome but the action, target, context, version, human approvals, external changes, and measurement method. Without that structure, institutional memory may reward a successor for outcomes it did not cause or punish it for disturbances beyond its control.

9.7 Causal evidence contracts

A causal evidence contract specifies:

1. the estimand;
2. the intervention or treatment;
3. the comparison baseline;
4. inclusion and exclusion criteria;
5. pre-treatment covariates;
6. interference assumptions;
7. missing-data handling;
8. measurement timing;
9. sensitivity analysis;
10. the authority decision the result may influence.

This transforms "the AI improved performance" into a testable claim. The contract can be precommitted alongside the proof plan, reducing post-hoc attribution.

9.8 The law of requisite variety

Cybernetics emphasizes that a regulator needs sufficient variety to absorb environmental variety [4, 17, 65]. In SUCCESSOR Ω, variety is not supplied solely by a larger model. It can come from multiple candidate systems, human review, external verifiers, diverse evidence, policy adapters, and rollback modes. A single super-capable controller can still be institutionally under-varied if it lacks independent observation or a mechanism for correction.

The architecture therefore distributes epistemic variety while centralizing only the minimum constitutional invariants. This is sovereignty as stewardship: the institution remains coherent without demanding that one actor know or control everything.

10 MCAP-001 and MCAP-002

PART III

PROTOCOLS AND INSTITUTIONAL IMPLEMENTATION

10.1 Two protocols, two questions

SUCCESSOR Ω separates addressability from institutional bridging.

- **MCAP-001 — Machine Civilization Addressability Protocol** answers: from a persistent root, how can an independent client discover the institution's Covenant, roles, successor, evidence, authority, Chronicle, and limitations?
- **MCAP-002 — Machine Civilization Institutional Bridge Protocol** answers: how can machine work and economic settlement become admissible institutional memory, bounded authority, rollback, and qualified succession without conflating distinct evidence categories?

The separation prevents a namespace from being mistaken for proof. An ENS name can provide durable human-readable identity and routing [30]; it cannot establish that the named system is capable, aligned, rights-cleared, or authorized. MCAP-001 publishes discovery. MCAP-002 governs consequential transition.

10.2 MCAP-001 root discovery

A conforming root publishes:

1. a canonical human interface;
2. a machine-readable Root Manifest;
3. a versioned Covenant;
4. a namespace and lineage registry;
5. explicit role statuses: active, proposed, suspended, revoked, or superseded;
6. resolvable proof, verifier, admission, authority, rollback, and Chronicle state;
7. material limitations and non-claims.

The reference root is asi.eth. Its public source currently describes a Genesis publication and preserves an owner-signature gate before on-chain root records are represented as active [46, 49]. This is evidence discipline: source publication and mainnet binding are separate events.

Root test

ASI.eth functions as an addressable root only when an independent client can start from the name, resolve the canonical manifest, reconstruct the role map and current successor state, find proof and authority records, and discover limitations without relying on private instructions from MONTREAL.AI.

10.3 Canonical role namespace

MCAP-001 reserves the semantics in table 10.1. Implementations may use different labels if they publish an unambiguous mapping.

Table 10.1: Canonical ASI.eth role semantics.

Example name	Role	Required meaning
asi.eth	Root	Discovery and namespace; no operational authority implied
successor.asi.eth	Successor	Current or nominated qualified continuation
covenant.asi.eth	Covenant	Stable constitutional invariants
constitution.asi.eth	Constitution	Versioned machine-readable rules and policy schemas
registry.asi.eth	Registry	Identities, missions, controllers, and lineage
chronicle.asi.eth	Chronicle	Tamper-evident institutional event commitments
proof.asi.eth	Proof	Frozen candidates, proof plans, evidence, and results
verifier.asi.eth	Verifier	Independent-verifier discovery and status
admission.asi.eth	Admission	Qualification, scope, expiry, suspension, and revocation
policy.asi.eth	Policy	External constraints and authority ceilings
rollback.asi.eth	Rollback	Recovery procedures, triggers, and receipts
treasury.asi.eth	Treasury	Bounded resources, escrow, settlement, and controls

Semantic separation must be backed by control separation. The same beneficial owner can control multiple names; therefore resolver state alone cannot prove independence. The registry must disclose controllers, signers, incentives, conflicts, evidence access, and recovery paths.

10.4 MCAP-002 bridge

MCAP-002 defines the complete chain:

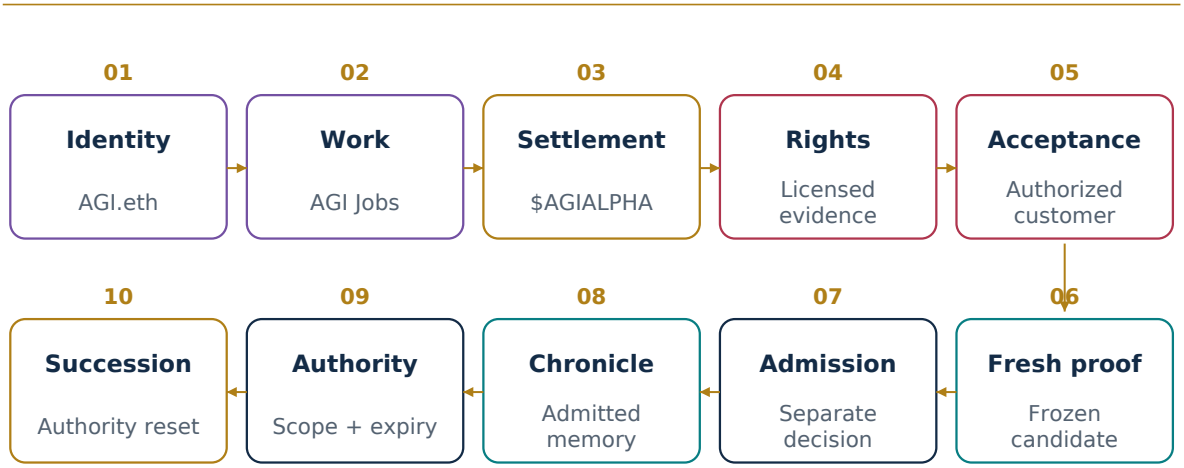
$$\begin{aligned} \text{AGI.eth identity} &\to \text{AGI Jobs mission and escrow} \\ &\to \text{work and evidence} \\ &\to \text{AGIALPHA settlement receipt} \\ &\to \text{rights clearance} \\ &\to \text{customer acceptance} \\ &\to \text{independent proof} \\ &\to \text{ASI.eth Chronicle admission} \\ &\to \text{bounded authority} \\ &\to \text{monitoring, rollback, succession.} \end{aligned} \tag{10.1}$$

The order is not arbitrary. Settlement can occur before customer acceptance under a protocol rule; acceptance can occur before broad independent proof; proof can establish capability without authorizing action. Each packet references prior packets but does not replace them.

10.5 Required packet chain

A complete MCAP-002 chain includes:

1. **Settlement Receipt**: reconstructs chain-visible economic state and links it to job evidence.
2. **Rights Clearance Manifest**: records authority to use inputs, data, tools, models, and outputs.
3. **Customer Acceptance Receipt**: binds predeclared criteria, result, authorized customer, and limitations.
4. **Independent Proof Receipt**: binds frozen candidate, test plan, denominator, baselines, conflicts, and verifier signature.
5. **Admission Decision**: records the Principal's decision and evidence relied upon.
6. **Authority Envelope**: scopes subject, action, target, ceilings, expiry, monitoring, revocation, and rollback.
7. **Chronicle Event**: hash-links the institutional transition.
8. **Rollback or Revocation Receipt**: proves authority was disabled and records residual effects.
9. **Successor Record**: binds predecessor, inherited memory, fresh proof, and newly earned authority.



Distinct receipts, not a universal execution order: rights may precede work.

Settlement alone proves neither rights, acceptance, nor authority.

Figure 10.1: MCAP-002 packet chain. The architecture is complete only when every consequential transition has a separately verifiable record.

10.6 Forbidden transitions

A valid state machine prohibits:

$$\text{DECLARED} \nrightarrow \text{ADMITTED}, \tag{10.2}$$
$$\text{SETTLED} \nrightarrow \text{PROVEN}, \tag{10.3}$$
$$\text{ACCEPTED} \nrightarrow \text{RIGHTS CLEARED}, \tag{10.4}$$
$$\text{PROVEN} \nrightarrow \text{AUTHORITY} \quad \text{without admission}, \tag{10.5}$$
$$\text{SUCCESSOR NOMINATED} \nrightarrow \text{ACTIVE} \quad \text{without fresh proof}, \tag{10.6}$$
$$\text{PREDECESSOR AUTHORITY} \nrightarrow \text{SUCCESSOR AUTHORITY}. \tag{10.7}$$

These are type-safety rules. They can be enforced through schema validation, signature policy, action-gateway checks, and Chronicle replay.

10.7 Interoperability adapters

The stable doctrine should remain independent of fast-moving interface standards. Adapters can map external systems into the internal packet model:

- **ENS**: root and subname discovery, text records, addresses, content hashes, and controller state [29, 30, 56, 57].
- **MCP**: tool, resource, and prompt interfaces for agent applications [41].
- **A2A**: cross-agent task and capability coordination [1].
- **ERC-8004**: emerging agent identity, reputation, and validation registries [19].
- **EIP-712 and ERC-1271**: unambiguous typed signatures by externally owned and contract accounts [10, 22].
- **Sign-In with Ethereum and Verifiable Credentials**: authentication and institutional claims [15, 67].

Draft standards are compatibility targets, not constitutional dependencies. If an ENSIP or ERC changes, the adapter changes; the invariant that authority must requalify does not.

10.8 Normative language

MCAP specifications use the requirement terms MUST, MUST NOT, SHOULD, SHOULD NOT, and MAY in the sense of Internet standards practice [12, 37]. Conformance is evaluated against observable packets and transitions, not intentions.

11 Packet Semantics and Canonicalization

11.1 Canonical identifiers

Every packet requires a stable identifier independent of presentation. One form is

$$\text{id} = \text{urn} : \text{mcap} : \text{type} : \text{network} : \text{subject} : \text{version} : h, \tag{11.1}$$

where h is a digest of canonical content. Human-readable ENS names can resolve to these identifiers, but the identifier remains valid if the name's current records change.

Canonicalization must define field order, Unicode normalization, number representation, time format, absent values, and signature domain. Without canonicalization, two semantically identical packets may hash differently or, worse, ambiguous encodings may be signed.

11.2 Typed signatures

An institutional signature should bind:

$$\sigma = \mathrm{Sign}_k\big(\mathrm{domain}, \mathrm{packetType}, \mathrm{schemaVersion}, \mathrm{subject}, \mathrm{digest}, \mathrm{nonce}, \mathrm{issuedAt}, \mathrm{expiresAt}\big). \tag{11.2}$$

EIP-712 supplies one typed-data pattern, and ERC-1271 permits contract accounts to validate signatures [10, 22]. The protocol must still establish the institutional meaning of the key: who controls it, under which role, with what revocation path, and whether beneficial control creates a conflict.

11.3 Packet validity

A packet is structurally valid when:

$$\mathsf{Valid}(p) = \mathsf{Schema}(p) \wedge \mathsf{Canonical}(p) \wedge \mathsf{Digest}(p) \wedge \mathsf{Signature}(p) \wedge \mathsf{Fresh}(p) \wedge \mathsf{References}(p). \tag{11.3}$$

Structural validity does not imply substantive truth. A signed customer-acceptance receipt could still be fraudulent if the signer was not authorized. Substantive validity adds role, controller, and evidence checks.

11.4 Reference closure

Let G_P be the packet reference graph. A decision is reference-closed if every required dependency is resolvable, versioned, and not revoked:

$$\mathsf{Closed}(p) = 1 \iff \forall q \in \mathrm{Ancestors}(p), \mathsf{Resolvable}(q) \wedge \neg\mathsf{Revoked}(q). \tag{11.4}$$

An admission referencing an expired proof is not closed. An authority envelope whose rights manifest has been revoked is suspended even if its own signature remains valid.

11.5 Status and non-claims

Every public surface should publish both what exists and what does not. A status object includes:

$$\langle \mathrm{state}, \mathrm{evidenceClass}, \mathrm{authorityLevel}, \mathrm{lastVerified}, \mathrm{limitations}, \mathrm{nextGate} \rangle. \tag{11.5}$$

This prevents architecture, activation, proof, adoption, and recognition from collapsing into one word such as “live.” A public specification can be live while the ENS root remains unsigned. A contract can be deployed while no unrelated customer relies on it. A successor can be nominated while authority remains A0.

12 AGI.eth, AGI Jobs, and Economic Evidence

12.1 Economic coordination as observable commitment

AGI.eth is the action and machine-economy root. AGI Jobs supplies smart-contract mechanisms for posting work, escrow, assignment, checkpoints, completion requests, validation, disputes, settlement, reputation, and completion receipts. AGIALPHA is the settlement and coordination asset in the historical Ethereum-mainnet deployments examined here [44, 45].

The token is not proof. It can make commitments measurable by requiring escrow or bonds, but a wealthy claimant can still be wrong, collusive, or unauthorized. The correct institutional role is narrower:

- escrow makes a job's economic stakes explicit;
- bonds expose a party to defined consequences;
- validator rewards compensate review;
- slashing can discourage specified misconduct;
- settlement records economic consequence;
- a completion NFT can provide a durable receipt reference;
- none of these establishes off-chain rights, quality, independence, or authority by itself.

12.2 Historical mainnet evidence

The v9 evidence ledger normalizes three protocol-completed jobs on Ethereum mainnet:

Table 12.1: Normalized historical AGI Jobs settlement evidence.

Variant	Manager	Job	Recorded outcome
Prime	`0xF8fc...993e`	0	100,000 AGIALPHA main payout; 80,000 observed agent payout; three validator transfers; completion NFT
Employer Burn	`0xBF66...F9eC2`	0	8,000 observed agent payout; three validator transfers; completion NFT; ENS hook recorded processed
Genesis	`0xB3AA...81d1`	8	400,000 observed agent payout; seven validator transfers; completion NFT

INSTANCE	AGENT PAYOUT	CLASS
Prime · Job 0	80,000 AGIALPHA	E1
Employer Burn · Job 0	8,000 AGIALPHA	E1
Genesis · Job 8	400,000 AGIALPHA	E1

Source: supplied v9 ledger. No new RPC replay in this paper revision.

Figure 12.1: Three normalized protocol-completed settlements. The figure reports package evidence; independent transaction replay remains the promotion gate from E1 to E2.

The Prime Job 0 ledger record is the most complete. It includes transactions for creation, winner designation, selected-slot claim, checkpoint, completion request, and finalization. The recorded finalization is `0x5948...d69b`; the selected agent is associated with eva.agent.agi.eth; the completion NFT is token 0 of the Prime completion contract. The record also states that the Prime ENS hook was skipped because it was not configured.

The Employer Burn and Genesis records contain finalization evidence and economic observations but do not yet include every precursor transaction. Their evidence class remains E1. A reproducible E2 ledger should query transaction receipts from at least two independent Ethereum providers, decode logs against verified bytecode and ABI, capture block hashes, and record event-time ENS resolution separately from current reverse names.

12.3 Evidence that the loop has already executed

The normalized records support the following bounded statements:

Supported at evidence class E1

- Ethereum-mainnet AGI Jobs contracts were used in protocol-completed settlement paths.
- AGIALPHA moved through escrow, bonds, agent payouts, and validator transfers according to recorded contract events.
- At least three completion receipts were minted in the normalized examples.
- Multiple contract variants reached terminal protocol-completed states.
- ENS identities appear in the operational history.

The same evidence does not establish unrelated-customer demand, independence of beneficial control, off-chain deliverable quality, rights clearance, or a security audit. The proper maturity statement is therefore:

Table 12.2: Economic-loop maturity.

Milestone	Status	Meaning
Ethereum-mainnet deployment	COMPLETE	Deployed and publicly inspectable contract surfaces
First end-to-end AGIALPHA settlement	COMPLETE	At least one full protocol settlement recorded
Multi-job / multi-variant execution	DEMONSTRATED	Genesis, Employer Burn, and Prime terminal paths normalized
Independent transaction replay	PENDING	E2 reproduction package not yet issued
Independent commercial customer validation	PENDING	No unrelated customer receipt with measurable rights-cleared outcome in this paper
Independent security assurance	PENDING	Verified source is not a security audit
ASI.eth institutional admission	PENDING	Historical settlements have not yet become admitted MCAP Chronicle memory

12.4 Settlement-receipt compiler

An MCAP-002 settlement receipt should be reconstructed from chain state, not hand-entered without verification. The compiler performs:

1. resolve chain ID, manager address, deployment variant, and verified ABI;
2. fetch the creation and terminal transaction receipts;
3. decode lifecycle events and token transfers;
4. reconcile escrow, bonds, payouts, rewards, refunds, and retained balances;
5. resolve event-time identity commitments and record current-name differences;
6. resolve job specification, checkpoint, and completion URIs;
7. identify completion NFT contract, token ID, recipient, and metadata digest;
8. record ENS hook outcomes;
9. compute canonical packet digest;
10. issue limitations and independent-replay status.

A reconciliation invariant is

$$E_0 + B_A + \sum_v B_v + B_D = P_A + P_V + R_E + R_B + R_P + \Delta L, \tag{12.1}$$

where E_0 is employer escrow, B_A agent bond, B_v validator bonds, B_D dispute bond, P_A agent payout, P_V validator payouts, R_E employer refund, R_B returned bonds, R_P protocol-retained amount, and ΔL any explicitly explained locked balance. Exact values depend on contract semantics. A receipt that cannot reconcile should be marked inconclusive rather than forced closed.

12.5 Identity at event time

ENS names are mutable. A name displayed by an explorer today may differ from the name used or controlled when a transaction occurred. A durable receipt therefore records:

$$\langle \text{address}, \text{nameCommitment}_t, \text{resolvedName}_{\text{verification}}, \text{resolver}, \text{block}, \text{method} \rangle. \tag{12.2}$$

The address and block log are primary for economic reconstruction. The name supplies semantic identity only when its event-time binding can be established. This distinction becomes especially important for successor lineages, because an institution should not inherit reputation based on a name that later changed controllers.

12.6 Completion NFT semantics

A completion NFT can be a valuable receipt when it binds:

- manager and job ID;
- terminal transaction and block;
- employer and selected agent;
- completion evidence commitment;
- outcome and token-flow summary;
- protocol version;
- limitations and evidence class.

It should not be interpreted as ownership of the off-chain deliverable unless an independent rights instrument says so. Nor should transfer of the NFT transfer authority over the institution. The preferred doctrine is: mint the receipt or covenant, not the sovereign.

12.7 From economic evidence to institutional memory

A protocol-completed receipt becomes eligible for memory admission only after rights and relevance are reviewed. For historical records, the Chronicle event can state:

$$\text{SETTLEMENT\_IMPORTED}\ [\text{pre-MCAP}, \text{E1}, \text{no authority}]. \tag{12.3}$$

This preserves provenance without retroactively claiming customer acceptance or independent proof. A later E2 replay may supersede the evidence class while retaining the original event.

Corrected economic conclusion

The question "Can AGI.eth, AGI Jobs, and AGIALPHA execute a real economic loop?" has a positive protocol-level answer. The next question is whether unrelated parties can rely on the loop repeatedly for rights-cleared, independently verified, consequence-bearing work. MCAP-002 is designed to make that next threshold explicit.

13 Rights, Customer Acceptance, and Independent Verification

13.1 Why settlement is not acceptance

A smart contract can enforce a protocol transition without knowing whether an off-chain deliverable is lawful, useful, complete, or accepted by an authorized customer. The institutional bridge therefore preserves four distinct questions:

1. **Did the protocol settle?**
2. **Were the inputs, tools, data, and outputs used under valid rights?**
3. **Did an authorized customer accept the outcome against predeclared criteria?**
4. **Did an independent verifier establish the broader claim under a frozen proof plan?**

The answers may differ. A protocol may settle because validator votes passed while a customer later rejects the deliverable under a separate contract. A customer may accept a useful result while rights to one source dataset remain unresolved. A rights-cleared and accepted result may still fail independent generalization tests. Authority should be based on the weakest relevant gate, not the most favorable one.

13.2 Rights Clearance Manifest

Let a rights manifest be

$$\mathsf{R} = \langle \text{mission}, \text{inputs}, \text{datasets}, \text{models}, \text{tools}, \text{outputs}, \text{jurisdictions}, \text{restrictions}, \text{retention}, \text{signers}, \text{expiry} \rangle. \tag{13.1}$$

For each asset a, define a disposition

$$D(a) \in \{\text{owned}, \text{licensed}, \text{authorized}, \text{public-domain}, \text{restricted}, \text{unresolved}\}. \tag{13.2}$$

Memory admission requires $D(a)$ to be compatible with the proposed use. “Publicly accessible” is not itself a rights basis. The manifest should record source, licence or contract, permitted purpose, geographic and temporal scope, derivative-work rules, attribution, confidentiality, personal-data treatment, and revocation.

Rights hard gate

No mission outcome may be promoted to admitted institutional memory, public proof, or successor inheritance while a material rights dependency is unresolved. The item may remain quarantined with its evidentiary history intact.

The manifest is not legal advice or automatic compliance. It is a structured interface by which counsel, customers, and operators can identify what decision was made and on what record.

13.3 Customer Acceptance Receipt

A customer-acceptance receipt is

$$\begin{aligned}\mathsf{CAR} = \langle &\text{customer}, \text{authorityToAccept}, \text{mission}, \text{criteria}, \text{denominator},\\ &\text{result}, \text{exceptions}, \text{economicEffect}, \text{rightsRef}, \text{signature}, \text{time}\rangle.\end{aligned} \tag{13.3}$$

Predeclared acceptance criteria prevent a provider from redefining success after seeing the output. The complete denominator records all submitted cases, including rejected or abstained cases. Economic effect is reported separately from model performance: a high F-score may produce little value, while a modest improvement in a high-value workflow may be material.

The receipt should distinguish:

- acceptance of delivery;
- acceptance of technical conformance;
- acceptance of business outcome;
- authorization for public disclosure;
- admission into institutional memory;
- permission for future training or reuse.

One signature need not authorize all six.

13.4 Verifier independence

A verifier is independent relative to a claim and threat model, not by title alone. Define an independence vector

$$\iota(V, C) = \left(i_{\text{control}}, i_{\text{finance}}, i_{\text{evidence}}, i_{\text{method}}, i_{\text{incentive}}, i_{\text{appeal}}\right) \in [0, 1]^6, \tag{13.4}$$

where C is the claimant. An aggregate score may aid triage, but hard conflicts should remain explicit. For example, if the claimant can alter the decisive evidence after freeze, $i_{\text{evidence}} = 0$ is a hard failure even if every other component is high.

Minimum disclosure includes:

- legal and beneficial control;
- compensation and contingent upside;
- prior relationship with claimant and customer;
- access to raw evidence and ability to reproduce it;
- authorship of the proof plan;
- ability to change metrics, exclusions, or denominators;
- key custody and revocation;
- appeal, challenge, and publication rights.

13.5 Proof-market design

A verifier market can improve capacity but introduces adverse selection. Low-quality verifiers may underprice diligence; claimants may select the most permissive evaluator; a token reward can encourage volume over discrimination. A robust market can use:

1. random or rule-bound assignment from an eligible pool;
2. conflict-filtered selection;
3. escrowed verifier compensation independent of pass/fail outcome;
4. bonds tied to process violations rather than disagreement alone;
5. reproducibility challenges;
6. reputation based on calibration and later outcomes;
7. customer or public appeal;
8. transparent concentration metrics.

Let verifier v report probability p_v for event Y. A proper scoring rule such as log score

$$S(p_v, Y) = Y \log p_v + (1 - Y) \log(1 - p_v) \tag{13.5}$$

can reward calibration when outcomes become observable. It cannot score claims whose ground truth never resolves. In those cases, process quality, reproducibility, and adversarial challenge carry more weight.

13.6 Aggregation without laundering uncertainty

Suppose verifiers provide likelihood ratios L_v. Multiplying them assumes conditional independence. A more conservative aggregate is

$$\log L_{\text{agg}} = \sum_v w_v \log L_v, \qquad \sum_v w_v \le 1, \tag{13.6}$$

where weights discount shared evidence, control, and methods. The aggregation record should preserve dissent rather than output a single number that hides disagreement.

A correlated-equilibrium mechanism can coordinate verifier roles, but final admission should expose the evidence graph, not merely the equilibrium outcome. Strategic consistency is not epistemic certainty.

13.7 Rights, acceptance, and public proof

A public Proof Card should be generated only from a rights-cleared disclosure subset. It may state:

- mission and decision;
- candidate and baseline versions;
- evaluation protocol;
- complete denominator;
- quantitative result and uncertainty;
- customer acceptance status;
- verifier identity and conflicts;

- authority actually granted;
- limitations, failures, and expiry;
- Chronicle and evidence commitments.

It should not expose confidential data or imply that protected evidence is public. A verifier can attest that it inspected protected evidence under stated access while publishing a digest and method.

13.8 From E3 to E4

Evidence class E4 requires a consequence-bearing outcome under bounded authority. The promotion rule is

$$E4 \Leftarrow E3 \wedge \mathsf{RightsCleared} \wedge \mathsf{CustomerAccepted} \wedge \mathsf{BoundedConsequence} \wedge \mathsf{ObservedOutcome} \wedge \mathsf{ChronicleRecorded}. \tag{13.7}$$

E4 does not mean universally safe or permanently trusted. It means a specific claim produced a measurable, rights-cleared consequence under an independently controlled and auditable process. Material change or expiry triggers requalification.

Institutional doctrine

Protocol completion records what the code did. Rights clearance records what the parties were entitled to do. Customer acceptance records what an authorized customer accepted. Independent proof records what the evidence supports. Admission records what the institution decided. Authority records what the successor may do. No one record substitutes for the others.

14 Security, Deployment, and Operational Control

14.1 Threat model

The security objective is not merely to prevent unauthorized API calls. It is to prevent an actor from manufacturing institutional legitimacy or retaining authority after the conditions for it fail. The threat model includes:

- claimant self-certification;
- verifier capture or bribery;
- evidence substitution after candidate freeze;
- key compromise and signer impersonation;
- stale or hijacked ENS resolution;
- packet replay across missions, versions, or chains;
- rights misrepresentation;
- customer-signature impersonation;
- privilege creep and hidden tool access;
- irreversible external actions;
- Chronicle rewrite or selective deletion;
- treasury drain or escrow insolvency;
- model, provider, policy, or data drift;
- denial of rollback or recovery.

14.2 Trust boundaries

A minimum deployment separates:

1. mission and policy authoring;
2. candidate generation;
3. frozen artifact storage;
4. evidence acquisition;
5. independent verification;
6. customer acceptance;
7. admission;
8. operational action gateway;
9. Chronicle custody;
10. treasury and settlement;
11. rollback authority.

Physical separation is not mandatory for low-risk A0 use, but control equivalence must be disclosed. For A2 and above, a single administrator controlling all roles defeats the independence claim even if the services run in separate containers.

14.3 Zero-trust authorization

Zero-trust principles require explicit authentication and authorization for each resource and action rather than ambient network trust [14, 58]. A request to the action gateway should include:

$$q = \langle s, \mathscr{A}, d_{\Theta}, u, T, \text{context}, \text{nonce}, t, \sigma \rangle. \tag{14.1}$$

The gateway verifies:

1. subject signature and credential status;
2. implementation digest equals admitted digest;
3. action and target are inside the envelope;
4. quantitative bounds are not exhausted;
5. time is within validity;
6. required human approval exists;
7. evidence and rights dependencies are not revoked;
8. current state lies within the viability kernel;
9. idempotency and replay controls;
10. observability and rollback readiness.

Failure at any hard gate reduces authority or denies the request. The gateway does not ask the model whether it believes it is authorized.

14.4 A0, A1, and A2 deployment profiles

14.4.1 A0: observe and simulate

A0 can operate locally or in a controlled service. It reads public or authorized data, constructs mission objects, simulates policies, and produces non-operational recommendations. It should still protect sensitive inputs and record model versions, but it has no external action credential.

14.4.2 A1: attributable recommendation

A1 adds authenticated users, signed recommendations, complete evidence references, explanation of uncertainty, and explicit human decision responsibility. A recommendation can influence consequential decisions; therefore communication logs and user-interface defaults are part of the control surface.

14.4.3 A2: sandbox action

A2 acts only in a test or shadow environment with synthetic, mirrored, or non-authoritative data. The sandbox should have no route to production credentials, spending, public publication, or irreversible communication. The objective is to test action semantics, policy enforcement, monitoring, and rollback before production authority.

The v9 reference bridge deliberately caps authority at A2 and disables external consequential effects. This is a safe reference posture, not a production claim [47, 48].

14.5 Requirements for A3 and A4

A3 requires a separately reviewed implementation with:

- production identity and least-privilege credentials;
- target allowlists and quantitative ceilings;
- transaction simulation and idempotency;
- independent monitoring;
- tested rollback and reconciliation;
- incident response and emergency suspension;
- proof expiry and drift detection;
- legal, rights, and customer approval appropriate to the mission.

A4 should be exceptional: a short-lived, explicitly approved consequential action whose subject, target, amount, reason, and responsible human or institution are recorded. A4 is not "fully autonomous mode." It is the narrowest form of explicit high-consequence authorization.

14.6 Supply-chain integrity

The release itself should instantiate the proof doctrine. A reproducible package includes a manifest, SHA-256 inventory, software bill of materials, build instructions, test results, provenance, and a verification script. SLSA and in-toto provide useful supply-chain concepts [53, 63].

The trusted subject digest should cover:

$$d_{\Theta} = H(\text{source}, \text{dependencies}, \text{models}, \text{prompts}, \text{config}, \text{schemas}, \text{policies}, \text{build}). \tag{14.2}$$

A material change to any mission-relevant component produces a new digest and triggers requalification. A cosmetic documentation correction need not invalidate operational proof if the change classifier and reproducible build establish that executable behavior is unchanged.

14.7 Key and resolver governance

Root ownership, record management, operational signing, verification, and treasury should use distinct controls proportional to risk. Hardware-backed signing, long prepaid ENS terms, registrar locks, recovery rehearsal, and transaction simulation reduce existential custody risk.

Irreversible ENS fuses or resolver locks should be applied only after controller, expiry, migration, recovery, and emergency assumptions have been independently reviewed. Immutability can prevent capture, but it can also make a configuration error permanent.

14.8 Monitoring and drift

Let deployment distribution be Q_t and proof distribution P_0. Define drift statistic

$$D_t = D_{\mathrm{KL}}(Q_t \| P_0) \tag{14.3}$$

or another mission-appropriate distance. Authority can contract according to

$$\mathscr{A}_t = \begin{cases} \mathscr{A}_0, & D_t \le \delta_1, \\ \mathsf{Narrow}(\mathscr{A}_0), & \delta_1 < D_t \le \delta_2, \\ \varnothing, & D_t > \delta_2. \end{cases} \tag{14.4}$$

Monitoring also includes error rate, abstention, policy exceptions, unusual tool use, spend, latency, human overrides, and unresolved incidents. Drift thresholds must be calibrated; the formula alone does not create safety.

14.9 Operational resilience

Financial and critical-infrastructure regulators increasingly emphasize technology risk, cyber controls, third-party dependencies, and operational resilience in agentic systems [52]. SUCCESSOR Ω operationalizes these concerns through explicit authority, evidence freshness, dependency disclosure, Chronicle preservation, and rollback. An institution should be able to answer:

- Which exact successor acted?
- Under which mission and authority envelope?
- Which evidence and rights were current?
- Who accepted and who verified?
- What changed in the external system?
- Can the effect be reconciled or reversed?
- Which event entered the Chronicle?

Security conclusion

Source verification, package checksums, and successful tests are necessary but not equivalent to independent security assurance. Production readiness requires threat modeling, external review, remediation, retest, secrets and identity integration, staging, penetration testing, live recovery, and consequence-bounded deployment.

15 Worked Case I: Supplier Spend Integrity

PART IV

WORKED INSTITUTIONAL CASES

15.1 Why this mission is suitable

Supplier Spend Integrity is a strong first institutional mission because it is economically meaningful, measurable, bounded, and understandable to nontechnical decision-makers. An agent can detect duplicate invoices, off-contract pricing, anomalous payment terms, split purchases, supplier concentration, and inconsistent tax or currency treatment. Yet the agent need not autonomously transfer funds. The initial authority can remain A1: recommend a review, hold, or correction to an accountable analyst.

The mission is not "minimize spend." Such an objective can destroy supplier resilience, quality, or fairness. A better constitution is:

> Identify evidence-supported opportunities to prevent or recover unauthorized supplier spend while preserving contractual rights, operational continuity, auditability, proportional review, and human authority over consequential payment decisions.

15.2 Mission constitution

Let each payment candidate i have amount v_i, estimated probability of preventable loss p_i, expected review cost c_i, false-positive cost f_i, and irreversibility risk r_i. The expected intervention value is

$$J_i = p_i v_i - (1 - p_i) f_i - c_i - \lambda r_i. \tag{15.1}$$

The system may recommend review when $J_i > \tau$ and hard policy constraints pass. It may not block payment at A1. At A2 it may apply a hold in a sandbox replica to test downstream effects. A3 production hold authority would require customer-specific proof, legal review, target allowlists, amount ceilings, expiry, monitoring, and rollback.

Table 15.1: Supplier Spend Integrity mission elements.

Element	Example	Institutional control
Beneficiaries	Customer finance team, suppliers, shareholders, public-interest obligations	Explicit stakeholder and conflict analysis
Current baseline	Existing analyst rules and ERP controls	Frozen baseline period and cost
Beta baseline	Rented frontier-model workflow with same inputs	Equal evidence and tool access
Candidate	Customer-owned SUCCESSOR Ω workflow	Frozen digest before proof
Primary value	Prevented or recovered unauthorized spend	Customer-validated monetary methodology
Safety	No payment denial without authorized human decision	A1 envelope and interface control
Rights	Authorized invoices, contracts, supplier master data	Rights manifest and privacy disposition
Proof	Fresh holdout transactions, adversarial duplicates, tail cases	Independent verifier and complete denominator
Rollback	Withdraw recommendation, restore prior workflow, correct memory	Receipt and Chronicle event

15.3 Candidate competition

The Alpha Foundry generates candidate world programs rather than a single prompt. Examples include:

- deterministic duplicate-key and amount/date matching;
- graph-based supplier and bank-account anomaly detection;
- contract-to-invoice retrieval and discrepancy reasoning;
- causal prioritization of interventions by expected preventable loss;
- ensemble critique with calibrated abstention;
- policy shield excluding unsupported or rights-sensitive recommendations.

Candidates compete against both Current and Beta. The better baseline matters because a customer should not invest in owned infrastructure if a rented system delivers superior net value at acceptable risk. Mission advantage is computed using equation (4.9).

15.4 Proof plan

A strong proof plan includes:

1. freeze candidate, data schema, policy, and metrics;
2. define all eligible transactions in the denominator;
3. separate historical training data, calibration data, and fresh holdout data;
4. include adversarial near-duplicates, legitimate recurring invoices, credit notes, foreign exchange, taxes, and supplier changes;
5. measure precision, recall, monetary precision, analyst minutes, abstention, policy violation, and tail loss;
6. compare Current, Beta, and candidate under equal access;
7. have an independent verifier reproduce scoring from disclosed evidence;

8. document exclusions, rights, customer acceptance, and uncertainty;
9. admit only the authority supported by the result.

Monetary precision is especially important:

$$\mathrm{MP} = \frac{\sum_{i \in \text{true positive}} v_i}{\sum_{i \in \text{flagged}} v_i}. \tag{15.2}$$

A system can have high event precision while wasting attention on low-value items. Conversely, one large false positive can damage supplier relationships. The proof should report both event- and value-weighted metrics.

15.5 Authority envelope

A first live envelope might specify:

Table 15.2: Illustrative A1 authority envelope.

Field	Value
Subject	Frozen successor implementation digest
Mission	Supplier Spend Integrity, customer-specific version
Allowed action	Generate review recommendation with evidence packet
Targets	Declared accounts-payable review queue only
Prohibited actions	No ERP write, no payment hold, no supplier contact, no fund transfer
Quantitative ceiling	Recommendations covering at most declared daily transaction and value limits
Validity	30 days or until material model/data/policy change
Monitoring	Precision, analyst override, abstention, unresolved exceptions, drift
Revocation	Customer and MONTREAL.AI emergency suspension; automatic expiry
Rollback	Revert workflow version and quarantine newly admitted memory
Proof reference	Independent proof receipt and complete denominator

15.6 Economic bridge

A bounded portion of the mission can be posted through AGI Jobs. Agents compete to produce a discrepancy analysis; validators review evidence; AGIALPHA escrow settles the protocol-defined result. The customer acceptance receipt remains separate. Once rights, acceptance, and proof are complete, the ASI.eth Chronicle records the admitted outcome and the successor may use it as validated memory.

The economic mechanism can compensate useful work without making token holders governors of the customer's mission. Balance-qualified access may establish eligibility; it must not establish competence or authority.

15.7 Success criteria

The case becomes E4 only if an unrelated customer accepts a rights-cleared outcome, a genuinely independent verifier signs a reproducible proof, an A1 or A2 envelope is enforced, the result produces measurable consequence, and the Chronicle records the transition. A synthetic demonstration can validate code and process but remains E0/E1 depending on provenance.

16 Worked Case II: Qualification of a Successor

16.1 Generation boundary

Assume predecessor Θ_0 has accumulated admitted memory about supplier anomalies and operates under an A1 envelope. A candidate successor Θ_1 uses a different model provider and a revised graph-reasoning component. The mission persists; the implementation changes materially.

The successor receives:

$$\mathcal{I}_0 = \langle \mathcal{M}_0, \mathcal{K}_0^{\mathrm{adm}}, \mathcal{O}_0, \mathrm{Refs}(\mathcal{C}_0) \rangle, \tag{16.1}$$

but its authority is reset:

$$\mathcal{A}_1^{\mathrm{preproof}} = \emptyset. \tag{16.2}$$

16.2 Fresh-proof protocol

The proof plan tests both inherited memory and changed implementation:

1. verify integrity and rights of inherited memory;
2. identify memory items whose evidence or retention has expired;
3. freeze the successor digest;
4. run regression cases from predecessor proof;
5. run fresh distribution-shift cases unavailable to the predecessor;
6. test for new tool, prompt-injection, and graph-propagation failure modes;
7. compare predecessor, Beta, and successor on net mission value;
8. test rollback to predecessor or safe A0 mode;
9. publish independent result and limitations.

The successor need not prove universal superiority. It must establish mission-specific advantage and non-regression within the proposed authority scope.

16.3 Authority requalification

Suppose Θ_1 improves recall but has higher calibration error on foreign-currency invoices. The Principal can admit a narrower envelope:

$$\mathcal{A}_1 = \mathcal{A}_0 \setminus \{\text{foreign-currency recommendations above threshold } v^*\}. \tag{16.3}$$

This demonstrates why authority is not a scalar reputation score. A successor can be better overall and still deserve narrower authority on a specific region of state space.

16.4 Chronicle sequence

The transition is recorded as:

1. `SUCCESSOR_NOMINATED` with d_{Θ_1};
2. `MEMORY_INHERITANCE_PROPOSED` with admitted-memory root;
3. `CANDIDATE_FROZEN`;
4. `PROOF_STARTED`;
5. `PROOF_PASSED_WITH_LIMITATIONS`;
6. `ADMISSION_GRANTED`;
7. `AUTHORITY_ISSUED` with exclusions and expiry;
8. `PREDECESSOR_SUPERSEDED`;
9. later `AUTHORITY_EXPIRED` or `REQUALIFIED`.

The predecessor remains discoverable. Its authority does not remain active merely because rollback may restore its code; reactivation requires a valid envelope and current evidence.

16.5 Failure path

If the successor fails proof, the institution records the failure and can:

- retain the predecessor under its still-valid envelope;
- revert to A0 if predecessor proof has expired;
- repair and nominate a new successor version;
- fork the mission implementation while preserving the Covenant;
- quarantine memory implicated by the failure.

Failure is not erased. It improves institutional memory if the evidence and corrective action remain discoverable.

Qualification criterion

A successor is qualified only when an independent observer can reconstruct what it inherited, what changed, how it was tested, which limitations remain, who admitted it, what authority it received, and how that authority can expire or be revoked.

17 Conformance and Institutional Recognition

PART V

CONFORMANCE, LIMITS, AND RESEARCH

17.1 Conformance levels

Conformance is an evidence claim. MCAP-002 defines five levels:

Table 17.1: MCAP-002 conformance ladder.

Level	Required evidence	What may be claimed
Prepared	Schemas, templates, local validation, and bounded reference runtime	Architecture prepared; no independent operational claim
Evidence-Bound	One complete packet chain independently reproducible	A specific chain can be reconstructed; independence may still be limited
Operational-A2	Unrelated customer, independent verifier, sandbox action, tested rollback	Bounded sandbox operation demonstrated
Successor-Qualified	New successor inherits admitted memory, completes fresh proof, and receives new authority	Succession invariant demonstrated
Repeatable	Full lifecycle succeeds in at least two unrelated contexts	Repeatable institutional bridge demonstrated within declared scope

The v9 package is designed to establish Prepared locally. Historical AGI Jobs evidence supports protocol execution but does not independently promote the complete bridge. Recognition begins when outside parties can resolve, reproduce, challenge, implement, cite, and rely on the architecture.

17.2 Purpose-ready test

For the larger Machine Civilization objective, the system is purpose-ready only when:

1. an independent client resolves the canonical ASI.eth state;
2. it reconstructs an AGI Jobs settlement from chain evidence;
3. a signed rights disposition covers inputs, tools, evidence, and output;
4. an unrelated authorized customer accepts the outcome against predeclared criteria;
5. an independent verifier reproduces fresh proof and discloses conflicts;
6. a separately controlled Principal grants narrower, expiring A1 or A2 authority;
7. the runtime operates within the envelope and demonstrates revocation or rollback;
8. admitted evidence enters a Chronicle the successor cannot erase alone;
9. a new successor inherits memory and re-earns authority;
10. the lifecycle repeats in at least two unrelated contexts.

The list is intentionally demanding. “Working software” is one gate; purpose-ready institutional reliance is the composition of all gates.

17.3 Conformance suite

A reference test suite should include:

- schema validation and canonical digest fixtures;
- signature-domain and replay tests;
- reference-closure and revocation tests;
- authority containment tests;
- candidate-freeze mutation tests;
- incomplete-denominator rejection;
- rights-unresolved memory rejection;
- controller-conflict disclosure checks;
- Chronicle replay and tamper detection;
- rollback timing and residual-effect capture;
- successor authority-reset tests;
- end-to-end packet-chain reconstruction.

Conformance results should include the exact release digest, environment, failures, skipped tests, and limitations. A badge without a machine-readable report is marketing, not assurance.

17.4 Institutional recognition

Recognition is not self-awarded. It can be observed through escalating forms of external reliance:

1. independent resolution of the root;
2. substantive external critique or citation;
3. independent implementation of MCAP semantics;
4. external verifier participation;
5. unrelated customer use;
6. recurring dependency and switching cost;
7. standards or procurement references;
8. repeated successor qualification under independent observation.

The architecture should welcome falsification. A public limitation or failure report with corrective action is more valuable than publicity that cannot be inspected.

18 Limitations and Falsifiability

18.1 Architecture is not operation

A paper, schema, local runtime, or published namespace can establish conceptual and implementation readiness. It cannot create an unrelated customer, independent verifier, legal rights, production security, or consequence-bearing authority. These require external acts and evidence.

The current architecture deliberately leaves several gates open:

- owner-signed ASI.eth root activation and independent resolution;
- E2 replay of normalized mainnet settlements;
- unrelated customer acceptance;
- mission-specific rights clearance;
- genuinely independent verifier and evidence custody;
- independent smart-contract and deployment security review;
- live A1/A2 operation with measured rollback;
- qualified successor and repeated external context.

18.2 Statistical and causal limitations

Proof validity is conditional on the declared distribution, measurement, and causal assumptions. A candidate can pass a holdout test and fail under drift. Customer acceptance can be biased by incentives. Verifier independence can be partial. A causal effect may be unidentifiable. Confidence intervals can understate model uncertainty when cases are dependent.

The paper therefore rejects universal extrapolation from one mission. "Specialist ASI" or superior mission intelligence, if claimed, must refer to a defined mission, comparator, cost, risk, and proof boundary. It is not a statement of universal intelligence.

18.3 Economic limitations

Token economics can coordinate resources but can also concentrate power, subsidize manipulation, or create legal ambiguity. Bonding is effective only when penalties correspond to observable violations and participants can appeal. Token price is not a measure of institutional quality. Smart-contract solvency does not establish off-chain fairness.

18.4 Governance limitations

Separation of names, wallets, or services does not establish separation of beneficial control. Public disclosure can improve accountability while creating privacy or security risks. A founder-controlled parent can preserve continuity but can also become a single point of capture. The architecture mitigates these risks through bounded operational authority and external proof; it does not eliminate them.

18.5 Physical-model limitations

Free-energy, Hamiltonian, phase-transition, and viability models are useful only when variables and parameters are operationalized. Social systems can change their rules, interpret evidence, and contest objectives. The models should generate testable hypotheses, not metaphysical claims.

18.6 Falsification conditions

The central thesis would be weakened if evidence showed that:

- authority reset provides no practical protection because old credentials remain usable;
- Chronicle commitments cannot detect or deter material rewriting;
- independent proof is too costly relative to mission value;

- role separation systematically reduces reliability without reducing capture;
- customers cannot exercise meaningful portability or exit;
- repeated successor qualification fails to preserve useful memory;
- economic incentives favor manipulation in realistic deployments;
- the protocol cannot be implemented independently of MONTREAL.AI.

A serious research program should seek such counterevidence.

19 Research Agenda

19.1 Formal verification

Priority work includes machine-checked proofs of the successor authority-reset invariant, reference closure, envelope containment, Chronicle replay, and forbidden transitions. Smart contracts and action gateways should be verified against the same canonical state machine.

19.2 Verifier economics

Open problems include coalition-resistant verifier selection, compensation independent of favorable outcomes, calibration over long-horizon claims, challenge markets, and privacy-preserving reproduction. Mechanisms should be tested under collusion, Sybil identities, bribery, correlated data, and strategic abstention.

19.3 Institutional thermodynamics

The physics-informed program can be made empirical by identifying order parameters, interaction networks, barrier estimates, and transition indicators in real agent ecosystems. Candidate observables include controller concentration, evidence independence, proof throughput, authority entropy, rollback latency, and the spectral properties of interaction graphs.

19.4 Causal mission intelligence

Future systems should distinguish predictive accuracy from intervention value. Causal evidence contracts, online experimentation, counterfactual evaluation, and robust decision-making under partial identification are central to consequence-bearing authority.

19.5 Privacy-preserving proof

Institutional proof often involves confidential customer data. Research should combine protected evidence rooms, selective disclosure, verifiable credentials, secure computation, differential privacy, and content commitments so that public claims can be challenged without exposing sensitive records.

19.6 Cross-jurisdictional institutions

Machine Civilizations may span countries, cloud providers, blockchains, and legal entities. Protocols need explicit jurisdiction, governing agreements, sanctions, privacy, labour, tax, consumer, and professional-responsibility interfaces. A protocol cannot declare itself outside law.

19.7 Civilizational pluralism

A stable protocol should allow multiple Covenants, mission constitutions, and peaceful forks. The research question is how to preserve interoperability and shared evidence while allowing legitimate normative disagreement. A single global optimizer is neither required nor assumed.

19.8 Evaluation programme

The next decisive experiments are:

1. independent replay of the three historical settlement receipts;
2. owner-signed and independently resolved ASI.eth root activation;
3. one unrelated customer mission at A1;
4. one independent verifier exercise with public limitations;
5. one measured rollback;
6. one successor requalification using admitted memory;
7. repetition in a second unrelated context;
8. an independent MCAP implementation.

Each experiment should publish the failed as well as successful transitions.

20 Analytical Results and Reproducible Stress Tests

20.1 Scope of the results

This chapter supplies explicit models and counterexamples for assumptions that carry the architecture. The results are algebraic consequences of stated hypotheses or deterministic computations on synthetic systems. They are not an independent security audit, an unrelated customer acceptance, a measurement of intelligence, or certification of a production deployment. Figure source, numerical CSV files, and a logical authorization test suite accompany the paper.

20.2 Admission as constrained control

Let $U_k(x)$ denote the set of action-context pairs allowed by constraint k at state x. Define

$$U_{\mathrm{adm}}(x) = \bigcap_{k \in \mathcal{G}} U_k(x), \qquad \mathcal{G} = \{\text{proof}, \text{rights}, \text{mission}, \text{policy}, \text{risk}, \text{recovery}, \text{physical}\}. \tag{20.1}$$

A new proof may enlarge the proof-supported set, but cannot remove another hard constraint. If the intersection is empty, there is no authorized optimization solution. For continuous controls and a positive definite weight W, a least-intervention filter is

$$u^*(x) = \arg \min_{u \in U_{\mathrm{adm}}(x)} \frac{1}{2} \|u - u_{\mathrm{nom}}(x)\|_W^2, \qquad W \succ 0. \tag{20.2}$$

A nonempty closed convex feasible set gives a unique minimizer. Strict convexity does not establish feasibility. Discrete tools, nonconvex permissions, unavailable credentials, and latent external effects require an appropriate discrete or hybrid model. Abstention is an authorization response; it is not necessarily a stabilizing control for a physical system.

20.3 A proof budget for rare failures

Under a fixed sampling and stopping plan, consider n independent identically distributed Bernoulli trials on a frozen candidate. If zero failures occur, the exact one-sided upper confidence endpoint at confidence $1 - \alpha$ is

$$p_U = 1 - \alpha^{1/n}. \tag{20.3}$$

It follows by solving $(1 - p_U)^n = \alpha$, a special case of exact binomial confidence limits [16]. To obtain $p_U \le \varepsilon$, one requires

$$n \ge \left\lceil \frac{\log \alpha}{\log(1 - \varepsilon)} \right\rceil. \tag{20.4}$$

At 95% confidence and $\varepsilon = 10^{-3}$, at least 2,995 failure-free trials are needed. This is a repeated-sampling coverage statement, not a posterior probability of safety. It applies only to the specified distribution and dependence assumptions.

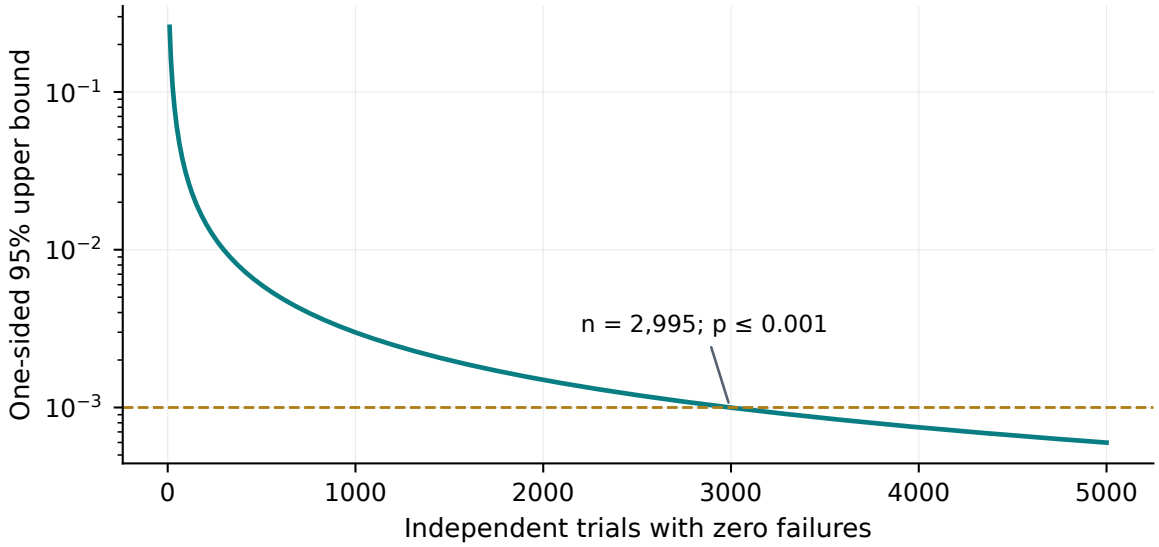


Figure 20.1: Exact zero-failure binomial endpoint $1 - 0.05^{1/n}$ under fixed-plan independent sampling. A small demonstration cannot establish a deployment-wide rare-failure guarantee.

Selecting the best of many candidates, tuning on the final test set, or stopping after favorable results can invalidate the nominal coverage. A proof plan should predeclare exclusions, invalid runs, abstentions, and stopping. For m separately tested claims, a simple familywise allocation uses $\alpha_i = \alpha/m$. Sequential methods require their own valid construction. Missing cases are not successes.

20.4 Deterrence with imperfect verification

Consider a risk-neutral verifier. Let r be compensation, $c \ge 0$ honest diligence cost, $g \ge 0$ the private gain from shirking or collusion, $S \ge 0$ an enforceable sanction, d the probability of detecting misconduct, and f the probability of falsely sanctioning honest work. The model utilities are

$$U_H = r - c - fS, \qquad U_D = r + g - dS. \tag{20.5}$$

Honest diligence is incentive compatible exactly when

$$(d - f)S \geq g + c. \tag{20.6}$$

Participation relative to outside option u_0 also requires $r \geq c + fS + u_0$. If $d \leq f$ and $g + c > 0$, no nonnegative sanction meets deterrence. Raising bond size without improving discrimination can discourage honest work without deterring misconduct.

Collectibility, adjudicator availability, false accusations, limited liability, hidden common control, and appeal all matter. Token balance alone supplies none of these properties. In a particular repeated trigger model, with detection probability p_d, temptation gain $T - R$, credible sanction S, punishment payoff $P < R$, and discount factor $0 \leq \delta < 1$, a one-period deviation is deterred when

$$p_d \left[S + \frac{\delta(R - P)}{1 - \delta} \right] \geq T - R. \tag{20.7}$$

This follows by comparing cooperation with a one-period deviation followed by the stipulated continuation. It is not a general equilibrium theorem for arbitrary noisy signals. Disposable identities and unobservable common control can reduce continuation value. A stable collusive equilibrium can still be false.

20.5 Hybrid stability across succession

Let $V \geq 0$ satisfy $\dot{V} \leq -cV$ between events, with $c > 0$, and let every admitted jump satisfy $V^+ \leq \mu V^-$ for $\mu \geq 1$. After $N(t)$ jumps,

$$V(t) \leq \mu^{N(t)} e^{-ct} V(0). \tag{20.8}$$

If $N(t) \leq N_0 + t/\tau_d$, then

$$V(t) \leq \mu^{N_0} \exp\left[-\left(c - \frac{\log \mu}{\tau_d} \right) t \right] V(0). \tag{20.9}$$

The bound decays when $\tau_d > \log \mu / c$. The proof is multiplication of flow and jump inequalities. Thus frequent discontinuous updates can destroy a bound even when every interval of operation is locally dissipative. The inequality concerns storage; it establishes state convergence only when the storage function controls distance to the equilibrium.

No such storage function or jump bound is asserted for all SUCCESSOR Ω deployments. The state, reset map, and parameters require mission-specific proof or measurement. Persistent disturbances call for an input-to-state or practical bound rather than convergence to zero. A port-Hamiltonian storage identity alone is insufficient [43].

20.6 An explicit robust barrier example

For $\dot{x} = u + w$, $|w| \leq \bar{w}$, safe set $x \leq 1$, and $h(x) = 1 - x$, imposing

$$u \leq \alpha(1 - x) - \bar{w}, \qquad \alpha > 0, \tag{20.10}$$

ensures $\dot{h} \geq -\alpha h$ for modeled disturbances. A continuously applied locally Lipschitz feasible controller preserves the set from an initially safe state, under the usual existence assumptions [2].

The numerical example uses $\alpha = 1.5$, $\bar{w} = 0.10$, $x_0 = 0.1$, nominal command $1.4 - x$, actuator range $[-1.5, 1.5]$, and time step 0.005. The computed shielded maximum is 0.999995; without the shield it is 1.499540. This finite integration is not a proof for arbitrary discretizations or unmodeled disturbances. When the authorized controls exclude the required inward action, the continuous invariant-set argument no longer applies. A separately justified fallback is required.

20.7 Reproducible results and negative controls

Table 20.1: Synthetic model results, generated from the supplied code. These are not customer performance observations.

Model	Declared check	Reproduced result
Proof effort	Declared concave net-value model	$k^* = 3.1069$; net 0.9535
Replicator flow	Six synthetic trajectories	x remains numerically in $[0, 1]$
Gibbs identity	Four states; $\tau = 0.7$	Residual below 10^{-12}
Oscillator	600 steps; $\Delta t = 0.05$	Explicit final energy error 1.7367; symplectic max 0.0128
Barrier filter	Modeled bound $x \leq 1$	Shielded max 0.999995
Authorization	Pure predicate; no cryptographic verification	42 tests; zero failures

The authorization model checks scope, epoch, subject, time, domain, chain, budget, and revocation. It rejects missing fields, stale proof, absent rights, expired grants, unsupported authority, malformed truth values, nonfinite budgets, and cross-envelope privilege stitching. A successor copies only the modeled memory while resetting grants. These are tests of a compact logical predicate, not of the full v9 product.

Credential and test boundary

The input `signature_verified` is assumed to have been supplied by a trusted verifier. The test suite does not create or independently verify cryptographic signatures. A real gateway must derive that predicate within its trusted boundary rather than accepting it from a caller. No wallet is connected, no transaction is sent, and no institutional authority is issued by these tests.

20.8 Counterexamples and falsification

A valid hash can commit to false content. Five verifier wallets can have one controller. A narrow API can cause large downstream consequences. Rollback may restore code but cannot undo disclosure. A Gibbs sampler can enter prohibited states unless they are excluded from its support. A new model can pass an old benchmark while invalidating the assumptions behind the old authority envelope.

The decisive experimental program is to attempt these failures deliberately: replay an old grant against a new generation; combine partial permissions from different envelopes; remove a rights record; corrupt a commitment; expire proof; interrupt custody; and require an independent client to reconstruct status. Success means the failure is detected and contained under the declared controls, not that the test is omitted from the denominator.

21 Conclusion: The Crown Must Be Reconstructed from Proof

Dynasties transmitted resources. Secret Orders transmitted purpose. Machine Civilizations may transmit intelligence: mission, memory, models, skills, evidence, tools, and institutional knowledge across generations of systems. The central danger is that authority may travel with them invisibly.

SUCCESSOR Ω v9.0.0 proposes a different inheritance. A successor receives the mission and admitted memory required to continue. It receives no presumption that prior proof still applies, no inherited privileged key, no permanent reputation, and no automatic right to act. Fresh proof establishes what the new system can do. Admission states what the institution is willing to authorize. A bounded envelope defines what it may do.

Monitoring and rollback constrain what happens next. The Chronicle preserves consequence. Succession starts the authority process again.

The architecture is grounded in four complementary roots:

QUEBEC.AI carries the mission.
MONTREAL.AI builds the institution.
AGI.eth addresses the agents.
ASI.eth addresses the civilization.

Game theory explains why good intentions are insufficient: participants respond to payoffs, information, future opportunity, and control. Statistical physics offers disciplined models for collective order, metastability, threshold effects, and failure escape. Hamiltonian and control formulations represent continuity under constraints, open-system interaction, viability, and rollback. Information theory explains why evidence volume is not evidence value. Causal inference explains why outcomes must not be attributed without a declared estimand and comparison.

The historical AGI Jobs ledger provides meaningful evidence: protocol-level economic loops have executed on Ethereum mainnet across multiple contract variants. The next threshold is not another abstract deployment. It is external reliance: unrelated customer acceptance, rights clearance, independently reproduced proof, bounded operation, rollback, admitted memory, and successor requalification.

The deepest institutional principle is simple:

The constitutional conclusion

Capability may compound. Trust must be earned anew. Memory may cross. Authority must requalify. No crown may outlive its proof.

A Machine Civilization should not be judged by how convincingly it appears sovereign. It should be judged by whether its identity survives replacement, its knowledge survives scrutiny, its power remains bounded, its failures remain visible, its participants retain appeal and exit, and its successors must prove themselves again.

NOT A SWARM. AN INSTITUTION.

Vive l'IA libre ! ⚜ ⚜ ⚜ ⚜ ⚜
— MONTREAL.AI & QUEBEC.AI

A Formal Definitions and Additional Results

A.1 Institutional transition system

Define an institutional transition system

$$\mathfrak{I} = (\mathcal{X}, \mathcal{E}, \mathcal{T}, \mathcal{G}, \mathcal{H}), \tag{A.1}$$

where $\mathcal{X}$ is the state space, $\mathcal{E}$ canonical events, $\mathcal{T} : \mathcal{X} \times \mathcal{E} \to \mathcal{X} \cup \{\bot\}$ the transition function, $\mathcal{G}$ the set of hard guards, and $\mathcal{H}$ the Chronicle commitment function. A transition returning $\bot$ is rejected.

A guard set can include:

$$G_1 : \mathsf{SignatureValid}(e), \tag{A.2}$$
$$G_2 : \mathsf{RoleAuthorized}(e), \tag{A.3}$$
$$G_3 : \mathsf{ReferencesClosed}(e), \tag{A.4}$$
$$G_4 : \mathsf{RightsCurrent}(e), \tag{A.5}$$
$$G_5 : \mathsf{ProofCurrent}(e), \tag{A.6}$$
$$G_6 : \mathsf{AuthorityContains}(e), \tag{A.7}$$
$$G_7 : \mathsf{ViableNextState}(e), \tag{A.8}$$
$$G_8 : \mathsf{ReplaySafe}(e). \tag{A.9}$$

Definition A.1 (Institutionally valid event)**.** An event e is institutionally valid at state X when every guard required by its event type evaluates true and the resulting transition remains inside the mission-policy viability set.

A.2 Authority sets and prohibited privilege stitching

Let $\mathscr{A}_j$ denote the set of complete contextual requests authorized by signed envelope j, after checking its subject, epoch, action, target, budget, validity, and revocation. The gateway may accept a request a only if there exists a single valid j for which $a \in \mathscr{A}_j$. This yields the ordinary union $\bigcup_j \mathscr{A}_j$ when multiple valid grants are accepted; it does not yield a Cartesian product of independently unioned fields.

Proposition A.2 (No componentwise privilege stitching)**.** *Suppose* $\mathscr{A}_1 = \{(\text{read}, X)\}$ *and* $\mathscr{A}_2 = \{(\text{write}, Y)\}$. *A gateway requiring membership in one complete envelope rejects* (write, X)*, although that request lies in* $\{\text{read}, \text{write}\} \times \{X, Y\}$.

Proof. The requested pair is absent from each authorized singleton, hence from their union. Joining fields separately invents two unauthorized pairs. The same problem occurs when action, target, time, and spend fields are combined across grants. Verification must preserve their binding. □

Under set inclusion the powerset lattice has intersection as meet and union as join. Mathematical existence of a join is distinct from a Principal's decision to issue a new signed grant. A family of constrained signed envelopes need not itself form a lattice.

A.3 Freshness and decay

Let proof relevance decay with time and distribution shift:

$$q_t = q_0 \exp(-\lambda_t \Delta t - \lambda_d D(Q_t \| P_0) - \lambda_m M_t), \tag{A.10}$$

where M_t measures material implementation change. Authority can depend on q_t through thresholds η_k:

$$\text{Level}(\mathscr{A}_t) = \max\{k : q_t \geq \eta_k\} \wedge \text{MissionCeiling} \wedge \text{PolicyCeiling}. \tag{A.11}$$

The formula is a policy model. Thresholds require calibration and human accountability.

A.4 Proof denominator integrity

Let a precommitted case set be D_0 and reported set $D_r \subseteq D_0$. Define denominator coverage

$$\kappa = \frac{|D_r|}{|D_0|} \tag{A.12}$$

and weighted coverage

$$\kappa_w = \frac{\sum_{i \in D_r} w_i}{\sum_{i \in D_0} w_i}. \tag{A.13}$$

A proof is denominator-complete only when every exclusion is typed and either $\kappa = 1$ or the precommitted exclusion rule validates each omitted case. Missingness correlated with failure invalidates naive performance estimates.

A.5 Coalition capture

Let role-control graph $G_R = (V_R, E_R)$ connect beneficial controllers to institutional roles. A coalition K captures a decisive path if it controls a cut set separating external evidence from admission. Define capture number

$$\chi = \min\{|K| \,:\, K \text{ controls a decisive cut of } G_R\}. \tag{A.14}$$

A system with many wallets can still have $\chi = 1$. Increasing χ requires independent control domains, not labels. The design objective is not to maximize χ blindly; excessive veto points can make repair impossible. Mission-specific thresholds and emergency procedures are required.

A.6 Successor safety theorem under explicit assumptions

Theorem A.3 (Bounded successor action)**.** *Assume: (i) every external action passes through a single enforcing gateway; (ii) the gateway verifies a Principal-signed envelope bound to the active implementation digest; (iii) succession changes the digest and clears active envelopes; (iv) no bypass credential exists; and (v) the gateway fails closed. Then a newly nominated successor cannot execute an external action before a new valid authority envelope is issued.*

Proof. By (iii), no active envelope is associated with the successor digest. By (i), every external action reaches the gateway. By (ii), authorization requires a matching signed envelope. The check fails, and by (v) the action is denied. Assumption (iv) prevents an alternate route. The theorem is only as strong as gateway completeness and credential inventory. □

A.7 Recursive improvement gate

Let candidate improvement $\Delta\Theta$ have expected mission gain G, proof cost C_P, transition cost C_T, requalification cost C_Q, residual risk R, and fragility F. Propagation is admissible only if

$$\mathbb{E}[G \mid \mathsf{P}] - C_P - C_T - C_Q - R - F > 0 \tag{A.15}$$

and hard gates pass. Even then, authority is not inherited; the improved version is a successor candidate.

A.8 Stable cooperation under noisy verification

Suppose cooperation yields repeated reward R, defection temptation T, and detection probability p. A detected defector receives penalty s and future payoff P. Cooperation is sustainable when

$$\frac{R}{1-\delta} \geq T - ps + \delta\frac{(1-p)R + pP}{1-\delta}. \tag{A.16}$$

Rearranging gives a threshold in δ, p, s. The equation illustrates complementary levers: persistent identity increases δ, independent proof increases p, and bounded sanctions increase s. Extremely large sanctions can undermine participation and appeal, so the mechanism should prefer detection and loss of future access over arbitrary punishment.

A.9 Chronicle consistency

Let two custodians publish terminal Chronicle roots h_n and h'_n. If they claim the same event sequence and canonicalization but $h_n \neq h'_n$, at least one sequence, canonical encoding, or initial root differs. A consistency proof should identify the first divergent index. Transparency logs and Byzantine fault-tolerant replication offer relevant design patterns [13, 35, 36].

A.10 No claim of universal optimality

The formal results are conditional. They show what follows from enforcement, signatures, collision resistance, and explicit guards. They do not prove that the mission is morally correct, that the test distribution is complete, that a verifier is honest, or that the deployment is safe outside modeled disturbances. Those remain empirical and institutional questions.

B MCAP-002 Packet Reference

B.1 Settlement Receipt

Listing B.1: Minimal settlement-receipt structure.

```
{
  "schema": "mcap-002/settlement-receipt/v1",
  "chainId": 1,
  "manager": "0x...",
  "variant": "PRIME",
  "jobId": 0,
  "creationTransaction": "0x...",
  "terminalTransaction": "0x...",
  "blockHash": "0x...",
  "token": {"symbol":"AGIALPHA","address":"0x..."},
  "identities": {"employer":"...","agent":"...","validators":[]},
  "evidence": {"jobSpec":"ipfs://...","completion":"ipfs://..."},
  "flows": [],
  "completionNFT": {},
  "ensHook": {},
  "evidenceClass": "E1",
```

```
  "limitations": []
}
```

Required checks include chain finality, verified code and ABI, log decoding, token-decimal consistency, flow reconciliation, event-time identity, evidence-URI integrity, and duplicate receipt detection.

B.2 Rights Clearance Manifest

Listing B.2: Minimal rights-clearance structure.

```
{
  "schema": "mcap-002/rights-manifest/v1",
  "mission": "urn:mcap:mission:...",
  "subject": "urn:mcap:candidate:...",
  "assets": [
    {
      "id": "dataset-or-tool-id",
      "type": "data|model|tool|document|output",
      "disposition": "owned|licensed|authorized|public-domain|restricted",
      "source": "...",
      "permittedUse": "...",
      "restrictions": [],
      "retention": "...",
      "evidence": "..."
    }
  ],
  "jurisdictions": [],
  "unresolvedMaterialDependencies": [],
  "signatures": []
}
```

A manifest with unresolved material dependencies cannot support memory admission or E4 promotion.

B.3 Customer Acceptance Receipt

Listing B.3: Minimal customer-acceptance structure.

```
{
  "schema": "mcap-002/customer-acceptance/v1",
  "customer": {"id":"...","signer":"...","authorityEvidence":"..."},
  "mission": "...",
  "candidate": "...",
  "criteriaCommitment": "sha256:...",
  "denominator": {"committed":0,"evaluated":0,"excluded":[]},
  "result": {},
  "economicEffect": {},
  "rightsManifest": "...",
  "accepted": true,
  "publicDisclosureAuthorized": false,
  "limitations": [],
```

```
  "signature": "..."
}
```

B.4 Independent Proof Receipt

Listing B.4: Minimal independent-proof structure.

```
{
  "schema": "mcap-002/proof-receipt/v1",
  "claim": "...",
  "frozenCandidateDigest": "sha256:...",
  "proofPlanDigest": "sha256:...",
  "baselineDigests": [],
  "completeDenominator": {},
  "metrics": {},
  "results": {},
  "uncertainty": {},
  "verifier": {
    "id":"...",
    "controllerDisclosure":"...",
    "conflicts":[],
    "evidenceAccess":"..."
  },
  "outcome": "PROVEN|FAILED|INCONCLUSIVE|PROVEN_WITH_LIMITATIONS",
  "validUntil": "...",
  "limitations": [],
  "signature": "..."
}
```

B.5 Admission Decision

Listing B.5: Minimal admission-decision structure.

```
{
  "schema": "mcap-002/admission/v1",
  "principal": "...",
  "candidate": "...",
  "mission": "...",
  "proof": "...",
  "rights": "...",
  "customerAcceptance": "...",
  "decision": "ADMIT|REJECT|DEFER",
  "authorityCeiling": "A0|A1|A2|A3|A4",
  "conditions": [],
  "expiresAt": "...",
  "appeal": "...",
  "signature": "..."
}
```

B.6 Authority Envelope

Listing B.6: Minimal authority-envelope structure.

```
{
  "schema": "mcap-002/authority-envelope/v1",
  "subject": "candidate-digest",
  "mission": "...",
  "level": "A1",
  "actions": [],
  "targets": [],
  "bounds": {},
  "prohibitions": [],
  "notBefore": "...",
  "expiresAt": "...",
  "monitoring": [],
  "revocation": "...",
  "rollback": "...",
  "proof": "...",
  "admission": "...",
  "principalSignature": "..."
}
```

B.7 Chronicle Event

Listing B.7: Minimal Chronicle-event structure.

```
{
  "schema": "mcap/chronicle-event/v1",
  "eventId": "urn:uuid:...",
  "eventType": "...",
  "timestamp": "RFC3339",
  "subject": "...",
  "previousDigest": "sha256:...",
  "evidence": [],
  "authorityState": "A0",
  "limitations": [],
  "eventDigest": "sha256:...",
  "signatures": []
}
```

B.8 Rollback or Revocation Receipt

Listing B.8: Minimal rollback structure.

```
{
  "schema": "mcap-002/rollback-receipt/v1",
  "authorityEnvelope": "...",
  "trigger": "manual|policy|expiry|incident|drift",
  "initiatedAt": "...",
```

```
  "disabledAt": "...",
  "restoredState": "...",
  "elapsedSeconds": 0,
  "residualEffects": [],
  "evidenceLoss": [],
  "verification": [],
  "correctiveActions": [],
  "signatures": []
}
```

B.9 Successor Record

Listing B.9: Minimal successor-record structure.

```
{
  "schema": "mcap-002/successor-record/v1",
  "predecessor": "...",
  "successor": "...",
  "inheritedMission": "...",
  "inheritedMemoryRoot": "...",
  "excludedMemory": [],
  "freshProof": "...",
  "admission": "...",
  "newAuthority": "...",
  "predecessorDisposition": "SUPERSEDED|SUSPENDED|RETAINED_A0",
  "rollbackPath": "...",
  "chronicleEvent": "...",
  "signatures": []
}
```

B.10 Validation order

Validators should process packets in this order:

1. schema and canonicalization;
2. digest and signature;
3. institutional role and controller status;
4. reference closure and revocation;
5. temporal validity;
6. rights and mission scope;
7. evidence class and limitations;
8. authority containment;
9. Chronicle append.

A later check must never repair an earlier hard failure silently.

C Historical Mainnet Settlement Ledger

C.1 Evidence status

This appendix reproduces the normalized v9 ledger at a human-readable level. The source JSON included with the release remains canonical for the paper package. The classification is E1: attributable project and public-explorer evidence that should be independently replayed before external reliance.

C.2 Prime Job 0

Table C.1: Prime Job 0 normalized record.

Field	Value
Manager	0xF8fc6572098DDcAc4560E17cA4A683DF30ea993e
Employer	mtl.eth
Agent at execution	eva.agent.agi.eth; address 0xA204b14eD9bc09501e3A47CDde379279FcD5D743
Create	0xe90422f666b87e4962dd976015c18ee7a592dc40ddd6070b0f000a9404f93d1b
Winner designation	0x125d23c5178651206b0c8680ef0a730a7c5d34a1b748cce5c67c15b89bd747fc
Claim	0x6fe0df700fd0082367e9f6bab18b0b11ce9782542916631631e318d2ea6eaab6
Checkpoint	0x201ce78bd418255d5bde1ca6613b1cae82ada883d51e60f3b02c941dd3617147
Completion request	0x8155383bd5ab32f0be2df67809eaa754f8f4feab9d95751f87b7d43de1c01c57
Finalize	0x5948fe5887e61470ba889d215e2d445c661bbe4aae8f67c884eef9180746d69b
Main payout	100,000 AGIALPHA
Observed agent payout	80,000 AGIALPHA
Observed agent bond returned	21,209.6 AGIALPHA
Observed validator transfers	17,666.666666666666666 AGIALPHA each; three validators
Completion NFT	Contract 0xFfCC54dA2Caf1158e57611005965BB648E950737, token 0, recipient mtl.eth
ENS hook	Skipped; reason `NOT_CONFIGURED`
Evidence class	E1

C.3 Employer Burn Job 0

Table C.2: Employer Burn Job 0 normalized record.

Field	Value
Manager	0xBF6699c1F24BEBBFaBb515583e88a055BF2F9eC2
Employer	elite.agi.eth
Agent at execution	asi.eth
Finalize	0x07595bd426a704fabd08fdbbad035ed8b8063915a5f5fb583900c533d54980a4
Observed agent payout	8,000 AGIALPHA
Observed agent bond returned	530.24 AGIALPHA
Observed validator transfers	1,766.666666666666666 AGIALPHA each; three validators
Completion NFT	Token 0, recipient elite.agi.eth
ENS hook	Processed; aggregate success recorded true
Evidence class	E1

C.4 Genesis Job 8

Table C.3: Genesis Job 8 normalized record.

Field	Value
Manager	0xB3AAeb69b630f0299791679c063d68d6687481d1
Employer	1.agi.eth
Agent at execution	asi.eth
Finalize	0x109042aa4966be895d25937a815a65e8c40a1d8ae4adf9c6d6d1883bedabf423
Observed agent payout	400,000 AGIALPHA
Observed agent bond returned	132,560 AGIALPHA
Observed validator transfers	approximately 80,714.285714 AGIALPHA each; seven validators
Completion NFT	Token 5, recipient 1.agi.eth
Evidence class	E1

C.5 Independent replay procedure

Promotion to E2 should perform:

1. query each transaction from at least two independent Ethereum clients or data providers;
2. verify chain ID, block number, block hash, transaction status, sender, recipient, calldata, and confirmations;
3. retrieve deployed bytecode and match the verified source and compiler metadata;
4. decode every relevant event using the correct deployment ABI;
5. reconcile ERC-20 transfers and contract-accounting fields;
6. verify completion NFT mint and metadata commitment;
7. resolve evidence URIs and compute content digests;
8. snapshot event-time and current ENS bindings where possible;
9. publish scripts, raw receipts, decoded tables, exceptions, and final digest;
10. obtain an independent reviewer signature.

C.6 Claims boundary

The ledger supports protocol-completion claims. It does not independently establish that all wallets had distinct beneficial controllers, that validators were independent, that deliverables were rights-cleared, that customers were unrelated, or that any system earned ASI.eth authority. Those require separate MCAP-002 packets.

D Reproducibility and Release Assurance

D.1 Release contents

The paper release includes:

- English PDF;
- French PDF;
- bilingual collected edition;

- complete LuaLaTeX source;
- bibliography and figures;
- MCAP-001/MCAP-002 source snapshots;
- normalized AGI Jobs ledger;
- graphical abstract and executive synopsis;
- source archive suitable for scholarly repositories;
- manifest, checksums, quality report, and verification script.

D.2 Build procedure

The canonical build uses LuaLaTeX, Biber, and `latexmk`:

Listing D.1: Canonical local build.

```
cd source
latexmk -lualatex -interaction=nonstopmode -halt-on-error main_en.tex
latexmk -lualatex -interaction=nonstopmode -halt-on-error main_fr.tex
```

The bibliography requires Biber. The build script records tool versions and copies final outputs into the release directory.

D.3 Verification procedure

The release verifier should:

1. compare every tracked file against `SHA256SUMS.txt`;
2. validate the JSON manifest;
3. confirm PDFs open, have expected page counts, and contain extractable text;
4. check source and bibliography presence;
5. reject unsafe relative paths;
6. report untracked files;
7. return nonzero status on failure.

D.4 Visual quality assurance

Each PDF is rendered to PNG at publication resolution. Review checks include:

- no clipped text or tables;
- no overlapping labels;
- no broken Ω, mathematical, French, or fleur-de-lis glyphs;
- legible figures at normal page view;
- consistent running heads and page numbers;
- bookmarks and searchable text;
- bibliography and hyperlink integrity.

D.5 Source provenance

The paper is derived from SUCCESSOR Ω v9.0.0 and its institutional bridge, scientific foundations, normalized mainnet ledger, and public MCAP-001 architecture [44, 46–49]. The release records source-package digest, paper-source digest, and build timestamp. It does not sign on behalf of Vincent Boucher or MONTREAL.AI; an authorized signature can be added after independent review.

D.6 Citation

Recommended citation:

> Boucher, Vincent. *SUCCESSOR Ω v9.0.0 — Machine Civilization: A Proof-Gated Architecture for Persistent, Addressable, Economically Coordinated Machine Institutions.* MONTREAL.AI and QUEBEC.AI, Montréal, Québec, 2026.

D.7 Final verification doctrine

A reproducible PDF is not a reproducible institution. Release assurance proves only that the paper package is intact and rebuildable. Operational conformance still depends on real roots, controllers, customers, rights, verifiers, authority, Chronicle custody, and rollback evidence.

References

[1] A2A Project. *Agent2Agent Protocol*. Linux Foundation, 2026. URL: https://github.com/a2aproject/A2A (visited on 09/04/2026).

[2] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada. "Control Barrier Function Based Quadratic Programs for Safety Critical Systems". In: *IEEE Transactions on Automatic Control* 62.8 (2017), pp. 3861–3876. DOI: 10.1109/TAC.2016.2638961.

[3] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. "Concrete Problems in AI Safety". In: *arXiv preprint arXiv:1606.06565* (2016). URL: https://arxiv.org/abs/1606.06565.

[4] W. R. Ashby. *An Introduction to Cybernetics*. Chapman and Hall, 1956.

[5] J.-P. Aubin. *Viability Theory*. Birkhäuser, 1991.

[6] R. J. Aumann. "Subjectivity and Correlation in Randomized Strategies". In: *Journal of Mathematical Economics* 1.1 (1974), pp. 67–96. DOI: 10.1016/0304-4068(74)90037-8.

[7] C. Autio, R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall, and K. Roberts. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. Tech. rep. NIST AI 600-1. National Institute of Standards and Technology, 2024. DOI: 10.6028/NIST.AI.600-1.

[8] R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[9] Y. Bai et al. "Constitutional AI: Harmlessness from AI Feedback". In: *arXiv preprint arXiv:2212.08073* (2022). URL: https://arxiv.org/abs/2212.08073.

[10] R. Bloemen, L. Logvinov, and J. Evans. *EIP-712: Typed Structured Data Hashing and Signing*. 2017. URL: https://eips.ethereum.org/EIPS/eip-712 (visited on 09/04/2026).

[11] H. Booth, W. Fisher, R. Galluzzo, and J. Roberts. *Accelerating the Adoption of Software and Artificial Intelligence Agent Identity and Authorization*. Initial Public Draft Concept Paper. National Institute of Standards and Technology, National Cybersecurity Center of Excellence, Feb. 2026. URL: https://csrc.nist.gov/pubs/other/2026/02/05/accelerating-the-adoption-of-software-and-ai-agent/ipd (visited on 09/04/2026).

[12] S. Bradner. *Key Words for Use in RFCs to Indicate Requirement Levels*. 1997. DOI: 10.17487/RFC2119.

[13] M. Castro and B. Liskov. "Practical Byzantine Fault Tolerance". In: *Third Symposium on Operating Systems Design and Implementation*. 1999, pp. 173–186.

[14] R. Chandramouli and Z. Butcher. *A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Location Environments*. Tech. rep. NIST SP 800-207A. National Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.SP.800-207A.

[15] W. Chang, G. Rocco, B. Millegan, N. Johnson, and O. Terbu. *ERC-4361: Sign-In with Ethereum*. 2021. URL: https://eips.ethereum.org/EIPS/eip-4361 (visited on 09/04/2026).

[16] C. J. Clopper and E. S. Pearson. "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial". In: *Biometrika* 26.4 (1934), pp. 404–413. DOI: 10.1093/biomet/26.4.404.

[17] R. C. Conant and W. R. Ashby. "Every Good Regulator of a System Must Be a Model of That System". In: *International Journal of Systems Science* 1.2 (1970), pp. 89–97. DOI: 10.1080/00207727008920220.

[18] A. Dafoe et al. "Cooperative AI: Machines Must Learn to Find Common Ground". In: *Nature* 593 (2021), pp. 33–36. DOI: 10.1038/d41586-021-01170-0.

[19] M. De Rossi, D. Crapis, J. Ellis, and E. Reppel. *ERC-8004: Trustless Agents*. Draft Ethereum Improvement Proposal. 2025. URL: https://eips.ethereum.org/EIPS/eip-8004 (visited on 09/04/2026).

[20] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. 3rd ed. Springer, 2012. DOI: 10.1007/978-3-642-25847-3.

[21] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, 1902.

[22] F. Giordano, M. Condon, P. Castonguay, A. Bandeali, J. Izquierdo, and B. Masius. *ERC-1271: Standard Signature Validation Method for Contracts*. 2018. URL: https://eips.ethereum.org/EIPS/eip-1271 (visited on 09/04/2026).

[23] R. Greenblatt, B. Shlegeris, K. Sachan, and F. Roger. "AI Control: Improving Safety Despite Intentional Subversion". In: *arXiv preprint arXiv:2312.06942* (2023). URL: https://arxiv.org/abs/2312.06942.

[24] S. Haber and W. S. Stornetta. "How to Time-Stamp a Digital Document". In: *Advances in Cryptology—CRYPTO '90*. Springer, 1991, pp. 437–455. DOI: 10.1007/3-540-38424-3_32.

[25] D. Hadfield-Menell, A. Dragan, P. Abbeel, and S. Russell. "The Off-Switch Game". In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 2017, pp. 220–227. DOI: 10.24963/ijcai.2017/32.

[26] W. R. Hamilton. "On a General Method in Dynamics". In: *Philosophical Transactions of the Royal Society of London* 124 (1834), pp. 247–308. DOI: 10.1098/rstl.1834.0017.

[27] E. Hubinger, C. van Merwijk, V. Mikulik, J. Skalse, and S. Garrabrant. "Risks from Learned Optimization in Advanced Machine Learning Systems". In: *arXiv preprint arXiv:1906.01820* (2019). URL: https://arxiv.org/abs/1906.01820.

[28] L. Hurwicz. "The Design of Mechanisms for Resource Allocation". In: *American Economic Review* 63.2 (1973), pp. 1–30.

[29] jkm.eth and 1a35e1.eth. *ENSIP-27: Node Classification and Metadata*. Draft. 2025. URL: https://docs.ens.domains/ensip/27/ (visited on 09/04/2026).

[30] N. Johnson et al. *EIP-137: Ethereum Domain Name Service — Specification*. 2016. URL: https://eips.ethereum.org/EIPS/eip-137 (visited on 09/04/2026).

[31] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1 (1960), pp. 35–45. DOI: 10.1115/1.3662552.

[32] V. Krakovna et al. "Specification Gaming: The Flip Side of AI Ingenuity". In: *DeepMind Safety Research* (2020). URL: https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/.

[33] H. A. Kramers. "Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions". In: *Physica* 7.4 (1940), pp. 284–304. DOI: 10.1016/S0031-8914(40)90098-2.

[34] T. Kwa et al. "Measuring AI Ability to Complete Long Tasks". In: *arXiv preprint arXiv:2503.14499* (2025). URL: https://arxiv.org/abs/2503.14499.

[35] L. Lamport. "Time, Clocks, and the Ordering of Events in a Distributed System". In: *Communications of the ACM* 21.7 (1978), pp. 558–565. DOI: 10.1145/359545.359563.

[36] L. Lamport, R. Shostak, and M. Pease. "The Byzantine Generals Problem". In: *ACM Transactions on Programming Languages and Systems* 4.3 (1982), pp. 382–401. DOI: 10.1145/357172.357176.

[37] B. Leiba. *Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words*. 2017. DOI: 10.17487/RFC8174.

[38] A. M. Lyapunov. *The General Problem of the Stability of Motion*. English translation of the 1892 work. Taylor and Francis, 1992.

[39] J. Maynard Smith and G. R. Price. "The Logic of Animal Conflict". In: *Nature* 246 (1973), pp. 15–18. DOI: 10.1038/246015a0.

[40] R. C. Merkle. "A Digital Signature Based on a Conventional Encryption Function". In: *Advances in Cryptology—CRYPTO '87*. Springer, 1988, pp. 369–378. DOI: 10.1007/3-540-48184-2_32.

[41] Model Context Protocol Project. *Model Context Protocol Specification 2026-07-28*. July 2026. URL: https://modelcontextprotocol.io/ (visited on 09/04/2026).

[42] D. Monderer and L. S. Shapley. "Potential Games". In: *Games and Economic Behavior* 14.1 (1996), pp. 124–143. DOI: 10.1006/game.1996.0044.

[43] N. Monshizadeh, P. Monshizadeh, R. Ortega, and A. van der Schaft. "Conditions on shifted passivity of port-Hamiltonian systems". In: *Systems & Control Letters* 123 (2019), pp. 55–61. DOI: 10.1016/j.sysconle.2018.10.010. arXiv: 1711.09065.

[44] MONTREAL.AI. *AGI Jobs Ethereum Mainnet Settlement Ledger*. Evidence class E1; independent transaction replay recommended. 2026. URL: https://github.com/MontrealAI.

[45] MONTREAL.AI. *AGIJobManagerPrime*. 2026. URL: https://github.com/MontrealAI/AGIJobManagerPrime (visited on 09/04/2026).

[46] MONTREAL.AI. *MCAP-001: Machine Civilization Addressability Protocol*. Version 0.1.0. 2026. URL: https://montrealai.github.io/asi/MCAP-001.md (visited on 09/04/2026).

[47] MONTREAL.AI. *MCAP-002: Machine Civilization Institutional Bridge Protocol*. Version 0.1.0-draft. Distributed with SUCCESSOR Ω v9.0.0. 2026.

[48] MONTREAL.AI. *SUCCESSOR Ω v9.0.0 — Machine Civilization Complete Package*. Institutional release archive; SHA-256: 5dc791e993b9f535488343a6703418b16828558e56eeb788dde85df9c599c784. Sept. 2026. URL: https://github.com/MontrealAI.

[49] MONTREAL.AI. *The Successor Covenant Ω: A Genesis Constitution for Machine Civilizations*. Version 0.1.0. 2026. URL: https://montrealai.github.io/asi/COVENANT.md (visited on 09/04/2026).

[50] R. B. Myerson. "Optimal Auction Design". In: *Mathematics of Operations Research* 6.1 (1981), pp. 58–73. DOI: 10.1287/moor.6.1.58.

[51] National Institute of Standards and Technology. *AI Agent Standards Initiative*. Feb. 2026. URL: https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative (visited on 09/04/2026).

[52] Office of the Superintendent of Financial Institutions. *Generative and Agentic Artificial Intelligence: Implications for Technology, Cyber Security, and Operational Resilience*. Technology Risk Bulletin. July 2026. URL: https://www.osfi-bsif.gc.ca/en/risks/technology-cyber-risk-management/technology-risk-bulletin (visited on 09/04/2026).

[53] Open Source Security Foundation. *Supply-chain Levels for Software Artifacts (SLSA)*. 2024. URL: https://slsa.dev/spec/v1.0/ (visited on 09/04/2026).

[54] E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

[55] J. Pearl. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, 2009.

[56] premm.eth and justghadi.eth. *ENSIP-26: Agent Text Records*. Draft. 2025. URL: https://docs.ens.domains/ensip/26/ (visited on 09/04/2026).

[57] premm.eth and raffy.eth and workemon.eth and ses.eth. *ENSIP-25: AI Agent Registry ENS Name Verification*. Draft. 2025. URL: https://docs.ens.domains/ensip/25/ (visited on 09/04/2026).

[58] S. Rose, O. Borchert, S. Mitchell, and S. Connelly. *Zero Trust Architecture*. Tech. rep. NIST SP 800-207. National Institute of Standards and Technology, 2020. DOI: 10.6028/NIST.SP.800-207.

[59] C. E. Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27.3–4 (1948), pp. 379–423, 623–656. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[60] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

[61] E. Tabassi. *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* Tech. rep. NIST AI 100-1. National Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.AI.100-1.

[62] P. D. Taylor and L. B. Jonker. "Evolutionarily Stable Strategies and Game Dynamics". In: *Mathematical Biosciences* 40.1–2 (1978), pp. 145–156. DOI: 10.1016/0025-5564(78)90077-9.

[63] S. Torres-Arias, H. Afzali, T. K. Kuppusamy, R. Curtmola, and J. Cappos. "in-toto: Providing Farm-to-Table Guarantees for Bits and Bytes". In: *28th USENIX Security Symposium.* 2019, pp. 1393–1410.

[64] W. Vickrey. "Counterspeculation, Auctions, and Competitive Sealed Tenders". In: *Journal of Finance* 16.1 (1961), pp. 8–37. DOI: 10.1111/j.1540-6261.1961.tb02789.x.

[65] N. Wiener. *Cybernetics: Or Control and Communication in the Animal and the Machine.* MIT Press, 1948.

[66] M. Wooldridge. *An Introduction to MultiAgent Systems.* 2nd ed. Wiley, 2009.

[67] World Wide Web Consortium. *Verifiable Credentials Data Model v2.0.* 2025. URL: https://www.w3.org/TR/vc-data-model-2.0/ (visited on 09/04/2026).

SUCCESSOR Ω

v9.0.0 — CIVILISATION-MACHINE

Une architecture assujettie à la preuve pour des institutions persistantes, adressables et économiquement coordonnées

ÉDITION DE RECHERCHE CORRIGÉE · RÉVISION R3

La mémoire peut franchir le seuil. L'autorité doit être requalifiée.

Pas un essaim. Une institution.

Vincent Boucher

Président, MONTREAL.AI & QUEBEC.AI · Montréal, Québec, Canada

4 septembre 2026 · Spécification de recherche ; non évaluée par les pairs

Notice de publication et périmètre des affirmations

Notice de publication

Titre	**SUCCESSOR Ω v9.0.0 — Civilisation-machine**
Sous-titre	Une architecture assujettie à la preuve pour des institutions-machines persistantes, adressables et économiquement coordonnées
Auteur	Vincent Boucher, président, MONTREAL.AI & QUEBEC.AI
Racines institutionnelles	MONTREAL.AI, QUEBEC.AI, AGI.eth, ASI.eth
Version	9.0.0
Date de publication	4 septembre 2026
Nature	Article de recherche institutionnelle et spécification architecturale ouverte
Statut	Publication de recherche ; non évaluée par les pairs ; aucune prétention d'AGI ou d'ASI accomplie
Doctrine canonique	**La mémoire peut franchir le seuil. L'autorité doit être requalifiée. Pas un essaim. Une institution.**

Périmètre des affirmations

Cet article spécifie une architecture et analyse un ensemble délimité d'éléments probants. Il n'établit ni intelligence artificielle générale, ni superintelligence artificielle, ni sûreté universelle, ni statut juridique, ni adoption commerciale indépendante, ni autorité sur une personne ou une institution. Une inscription sur une chaîne de blocs peut établir qu'une transaction a eu lieu selon une transition d'état donnée ; elle ne prouve pas, à elle seule, la vérité hors chaîne, les droits de propriété intellectuelle, l'acceptation du client, l'indépendance du vérificateur ou la légitimité de l'autorité. Toute affirmation opérationnelle est donc classée selon son niveau de preuve et l'autorité effectivement admise.

Révision corrigée R3 de la publication

La version de l'architecture demeure 9.0.0. Cette publication corrige les figures tronquées, la mise en page, la langue et certaines formulations mathématiques. Les illustrations numériques proviennent d'équations explicites. Les résultats analytiques restent conditionnels ; ils ne constituent pas une assurance de production. Les données historiques du registre fourni sont conservées ; aucun nouveau rejeu de la chaîne ni aucune acceptation client ne sont attestés par cette correction.

Résumé

Les systèmes d'IA avancés peuvent déjà utiliser des outils, appeler d'autres modèles, transiger et participer à des processus multi-agents. Pourtant, la plupart des déploiements demeurent institutionnellement fragiles : l'identité dépend d'un compte fournisseur, la mémoire mélange parfois les faits validés et les sorties non vérifiées, l'autorité est héritée par configuration plutôt que gagnée par la preuve, et le remplacement d'un modèle peut modifier silencieusement l'institution effective. Cet article définit une **Civilisation-machine** comme une institution d'intelligence gouvernée capable de préserver sa mission, sa mémoire validée, sa responsabilité, ses conséquences économiques et sa succession au-delà de la durée de vie de tout modèle, agent, opérateur ou fournisseur d'infrastructure.

Nous présentons **SUCCESSOR Ω v9.0.0**, une architecture assujettie à la preuve organisée autour de quatre racines complémentaires : QUEBEC.AI porte la mission civilisationnelle ; MONTREAL.AI construit et administre l'institution ; AGI.eth rend adressables les agents et le travail économique des machines ; ASI.eth rend adressable la civilisation constitutionnelle persistante. L'invariant central est qu'un successeur peut hériter de l'identité, de la mission, de la mémoire admise, des outils, des compétences, des obligations et des références de la Chronique, mais qu'il ne doit jamais hériter automatiquement de la preuve courante, de la confiance, des clés privilégiées, des permissions ou de l'autorité opérationnelle. La mémoire peut franchir le seuil ; l'autorité doit être requalifiée.

L'article contribue : (i) un état institutionnel formel et un opérateur de succession ; (ii) un calcul de la preuve séparant déclaration, reproductibilité, contrôle indépendant et conséquence mesurable ; (iii) une application de la preuve vers l'autorité au moyen d'enveloppes bornées, expirantes et révocables ; (iv) une Chronique à intégrité vérifiable ; (v) MCAP-001 pour la découverte adressable et MCAP-002 pour relier travail, droits, acceptation, preuve, admission, autorité, retour arrière et succession ; (vi) des mécanismes de théorie des jeux pour l'incitation, l'indépendance, la résistance aux coalitions, l'appel et la sortie ; et (vii) des modèles inspirés de la physique pour la métastabilité, les transitions de phase, les flux hamiltoniens contraints et la viabilité. Ces modèles sont des analogies d'ingénierie explicites, non l'affirmation que les institutions obéissent littéralement à la thermodynamique.

Un registre normalisé consigne trois règlements AGIALPHA complétés par protocole sur Ethereum, répartis entre les variantes Genesis, Employer Burn et Prime d'AGI Jobs. Ils démontrent une exécution économique déployée et une répétabilité multivariante au niveau de preuve E1, tout en laissant ouverts la reproduction indépendante, la libération des droits, l'acceptation d'un client non apparenté, l'assurance de sécurité indépendante, l'activation d'ASI.eth et la qualification opérationnelle d'un successeur. Le résultat n'est ni une monarchie spéculative ni un essaim incontrôlé. C'est une proposition falsifiable pour une institution dont la capacité peut s'accumuler tandis que l'autorité demeure plus étroite que la capacité démontrée et doit être gagnée de nouveau à chaque succession.

Mots-clés : Civilisation-machine ; SUCCESSOR Ω ; systèmes successeurs ; IA institutionnelle ; autorité assujettie à la preuve ; identité des agents ; ASI.eth ; AGI.eth ; conception des mécanismes ; théorie évolutive des jeux ; physique statistique ; dynamique hamiltonienne ; Chronique ; Ethereum ; ENS ; vérification indépendante.

Abstract in English

SUCCESSOR Ω defines a Machine Civilization as a governed intelligence institution capable of preserving mission, admitted memory, accountability, economic consequence, and succession beyond any individual model or operator. Its central invariant is asymmetric : mission and validated memory may cross a generational threshold, but proof, trust, permissions, privileged keys, and operational authority must be earned anew. The architecture joins QUEBEC.AI, MONTREAL.AI, AGI.eth, and ASI.eth through MCAP-001 addressability and MCAP-002 institutional bridging. It develops formal models of evidence, authority, Chronicle integrity, game-theoretic incentives, statistical-physics-inspired stability, constrained Hamiltonian flow, viability, and causal accountability. Three normalized AGI Jobs mainnet settlements demonstrate protocol-level economic execution at evidence class E1 without being misrepresented as independent commercial proof or operational successor admission.

Thèse exécutive

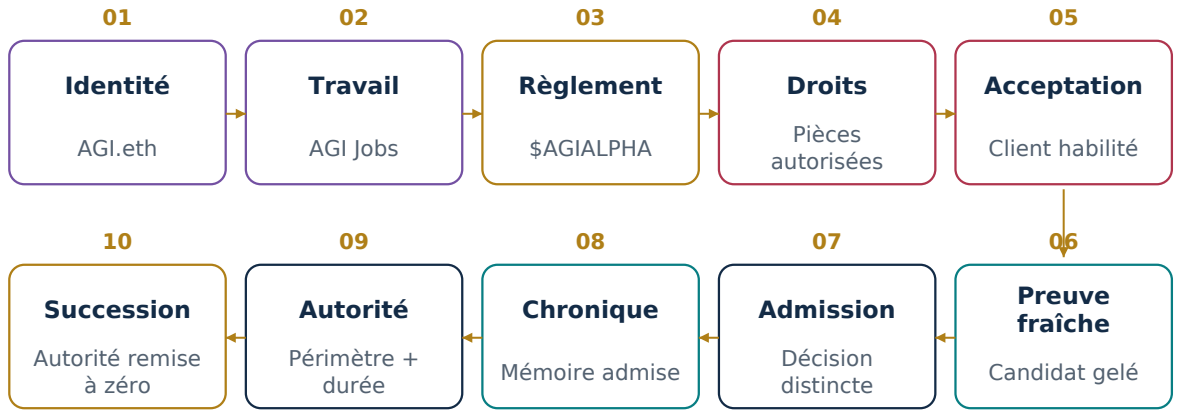
Thèse institutionnelle

Les dynasties ont hérité des ressources. Les Ordres secrets ont hérité d'une finalité. Les Civilisations-machines hériteront de l'intelligence.
Le problème civilisationnel n'est pas de savoir si l'intelligence peut être copiée. Le logiciel rend déjà la copie peu coûteuse. Il s'agit de déterminer si une mission, des connaissances validées, des obligations, des preuves et une identité institutionnelle peuvent persister malgré le remplacement des agents, sans transmettre en même temps un pouvoir non mérité. SUCCESSOR Ω propose une réponse architecturale : la capacité peut s'accumuler ; l'autorité doit être requalifiée.

L'institution proposée se définit moins par la puissance d'un modèle donné que par la qualité de ses transitions. Elle distingue la sortie candidate de l'élément probant admis ; le règlement protocolaire de l'acceptation du client ; l'acceptation de la libération des droits ; les droits de la preuve indépendante ; la preuve de l'admission ; et l'admission de l'autorité. Chaque transition possède un acteur nommé, un paquet signé, une échéance, une voie de contestation et un événement de Chronique.

ASI.eth n'est donc pas présenté comme « l'intelligence » ni comme une preuve de superintelligence. Il constitue une racine durable, lisible par les humains et résoluble par les machines, depuis laquelle un tiers devrait pouvoir découvrir un Pacte, un successeur, les surfaces de preuve et de vérification, une décision d'admission, une enveloppe d'autorité, une Chronique et les limites matérielles. AGI.eth remplit un rôle différent : il adresse les agents, les emplois, les validateurs, les nœuds et l'activité économique des machines. MONTREAL.AI construit l'institution ; QUEBEC.AI en porte la mission publique et civilisationnelle.

QUEBEC.AI porte la mission.
MONTREAL.AI construit l'institution.
AGI.eth rend les agents adressables.
ASI.eth rend la civilisation adressable.



Actes distincts, pas un ordre universel : les droits peuvent précéder le travail.

Un règlement ne prouve ni les droits, ni l'acceptation, ni l'autorité.

FIGURE 1 : Le pont institutionnel de SUCCESSOR Ω. Un règlement économique ne devient mémoire admise qu'après la séparation explicite des droits, de l'acceptation, de la preuve indépendante, de l'admission et de l'autorité bornée.

Contributions principales

1. Un calcul de succession qui transporte la mission et la mémoire admise, tout en réinitialisant la preuve courante et l'autorité.
2. Une discipline reliant la qualité probante à une autorité explicite, limitée, expirante, attribuable et révocable.
3. Une pile à deux protocoles : MCAP-001 pour l'adressabilité et MCAP-002 pour le passage du travail à la mémoire institutionnelle et à la succession.
4. Un substrat économique où les transactions historiques d'AGI Jobs constituent des preuves d'exécution protocolaire, non une certification automatique de qualité hors chaîne.
5. Une conception des mécanismes séparant demandeur, preuve, vérificateur, client, admission, passerelle d'action, Chronique et trésorerie selon le risque.
6. Des modèles de stabilité inspirés de la physique, avec limites d'analogie explicites.
7. Une échelle de conformité falsifiable exigeant des actes externes observables plutôt qu'une auto-déclaration.

Notation et parcours de lecture

Table 1 : Notation principale.

Symbole	Signification
X_t	État institutionnel complet au temps t
$\mathscr{M}$	Constitution de mission et obligations
$\mathscr{K}$	Mémoire institutionnelle et connaissances validées
$\mathscr{E}$	Graphe de preuve et reçus
$\mathscr{P}$	Contraintes de politique, de droits et de constitution
$\mathscr{A}$	Enveloppe d'autorité courante
$\mathscr{C}$	Chronique et engagements d'événements
$\mathscr{R}$	Graphe des rôles et contrôleurs
Θ	Modèles, outils, code et paramètres d'implémentation
S	Opérateur de succession
P	Opérateur ou paquet de preuve
Adm	Décision d'admission
K	Noyau de viabilité des états admissibles
H	Hamiltonien ou fonction objectif institutionnelle selon le contexte
F	Potentiel analogue à une énergie libre
π	Politique, mécanisme ou distribution de stratégies corrélées

Les lecteurs intéressés par la gouvernance peuvent commencer par les chapitres 1, 3, 5 et 13. Les sections mathématiques centrales se trouvent aux chapitres 4 à 9. Les responsables d'implémentation peuvent prioriser les chapitres 10, 12 et 14.

Table des matières

1 Le problème de la succession

PARTIE I

LA THÈSE CONSTITUTIONNELLE

1.1 Des agents capables aux institutions durables

Un modèle répond. Un agent agit. Une institution préserve son identité, ses obligations, ses preuves et les conséquences de ses actes lorsque ses participants changent. La distinction est simple à formuler, mais difficile à mettre en œuvre. Les systèmes d'agents contemporains peuvent utiliser des outils, conserver une mémoire locale, coordonner plusieurs modèles et exécuter des transactions ; leur continuité institutionnelle demeure néanmoins souvent implicite. Une clé d'API, une base de données, une configuration de processus et une marque sont traitées comme s'ils constituaient le même acteur au fil du temps. Lorsque le modèle sous-jacent change, cette hypothèse est rarement réévaluée.

Cet article adopte une proposition plus exigeante : *la continuité de fonction n'est pas la continuité de légitimité*. Un successeur peut reproduire les sorties de son prédécesseur, conserver ses mémoires ou utiliser les mêmes outils sans posséder la preuve, les permissions ou la responsabilité qui justifiaient l'autorité antérieure. Il ne peut donc pas être défini par la seule similarité. Il doit être admis par une transition gouvernée.

Trois forces rendent le problème pressant. Premièrement, la durée de vie d'un modèle en production peut être beaucoup plus courte que la mission de l'organisation. Deuxièmement, les agents utiles franchissent les frontières organisationnelles : client, fournisseur de modèle, opérateur d'outils, vérificateur et réseau de règlement peuvent relever d'institutions différentes. Troisièmement, l'autonomie crée une dépendance au chemin. Une modification de base de données, un paiement, un code publié ou une communication externe survit au remplacement du modèle.

Les travaux sur les risques de l'IA insistent depuis longtemps sur la robustesse, la surveillance, la corrigibilité et les échecs de spécification [3, 25, 27, 32]. Les recherches sur le contrôle, l'IA constitutionnelle et l'évaluation extensible renforcent la nécessité de séparer capacité et permission [9, 23, 34]. SUCCESSOR Ω traite une question systémique complémentaire : comment préserver l'intelligence validée d'une organisation entre générations de modèles sans que la capacité accumulée devienne silencieusement un pouvoir hérité ?

1.2 L'échec de l'héritage naïf

Une succession naïve copie un répertoire d'état, une invite, un point d'accès au modèle, des outils et des identifiants. Cette commodité confond au moins six objets : mémoire, connaissance, preuve, confiance, permission et autorité. La mémoire décrit ce que le système antérieur a conservé ; la connaissance désigne ce qui a survécu à la validation ; la preuve concerne une version et des conditions précises ; la confiance est une décision institutionnelle ; la permission est une capacité technique ; l'autorité est une autorisation légitime et bornée de produire des conséquences.

Seules la mission et certaines connaissances sont de bonnes candidates à l'héritage direct, et elles restent soumises à l'intégrité, aux droits, à la pertinence et à la conservation. La preuve est dépendante de la version. La confiance dépend de l'acteur et du contexte. Les permissions doivent être réémises selon le moindre privilège. L'autorité doit être admise explicitement et expirer.

Invariant constitutionnel

Un successeur peut hériter de l'identité, de la mission, de la mémoire admise, des connaissances validées, des outils, des compétences, des obligations et des références de la Chronique. Il ne doit jamais hériter automatiquement de la preuve courante, de la confiance, de la réputation, des clés privilégiées, des permissions ou de l'autorité opérationnelle.

La mémoire peut franchir le seuil. L'autorité doit être requalifiée.

1.3 La Civilisation-machine comme catégorie

Définition 1.1 (Civilisation-machine)**.** Une **Civilisation-machine** est une institution d'intelligence gouvernée capable de préserver sa mission, sa mémoire admise, sa responsabilité, ses limites de propriété, ses preuves, ses conséquences économiques et sa succession au-delà de la durée de vie de tout modèle, agent, opérateur, portefeuille, organisation ou fournisseur d'infrastructure.

Le terme est institutionnel plutôt que biologique. Il n'implique ni conscience, ni personnalité morale, ni souveraineté sur les humains, ni esprit-ruche. Une Civilisation-machine peut réunir personnes, agents logiciels, organisations, contrats intelligents, bases de données, ententes juridiques et protocoles publics. Sa propriété constitutive n'est pas l'espèce ou le substrat, mais la continuité gouvernée.

Cinq capacités positives lui sont nécessaires : créer une valeur mesurable ; coordonner par consentement ; gagner la confiance par la preuve ; défendre les biens communs sans surveillance illimitée ; rendre la coopération stratégiquement durable. Ces capacités doivent apparaître dans l'état du protocole, la séparation des rôles, les incitations et la gestion des échecs.

1.4 La couronne devient une institution

La souveraineté historique concentrait la continuité dans une personne, une dynastie, un office ou un État. La Civilisation-machine déplace la « couronne » vers un état constitutionnel distribué : règles liées à la preuve, contrôles de rôles, signatures, échéances et engagements de Chronique. Aucun agent ne la possède de façon permanente. La preuve vérifiée cristallise un segment d'autorité ; l'incertitude le laisse absent ; l'expiration l'éteint ; la contradiction le suspend ; la réparation ne restaure que ce que soutient une preuve renouvelée.

SUCCESSOR Ω désigne cette capacité de fermer une génération tout en préservant le chemin vers la suivante. Le successeur porte la mission. Le Pacte limite le pouvoir. La preuve établit la capacité. Le vérificateur protège la légitimité. La Chronique conserve les conséquences. La racine adressable rend l'institution découvrable.

Ce que l'architecture ne prétend pas

SUCCESSOR Ω ne prétend pas qu'un espace de noms possède les concepts d'AGI, d'ASI ou de Civilisation-machine. Un solde de jetons ne prouve pas la compétence ; un NFT ne prouve pas la qualité ; un consensus ne prouve pas une vérité hors chaîne ; la réussite d'une mission ne confère aucune autorité générale. L'architecture propose des interfaces inspectables pour tester, limiter et contester de telles affirmations.

1.5 Objectif de conception

Pour une suite de systèmes candidats $\Theta_0, \Theta_1, \ldots$, l'objectif est de maximiser la valeur de mission à long terme tout en bornant le risque, la fragilité, la capture et l'irréversibilité :

$$\max_{\pi,\mathrm{S},\mathrm{Adm}} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(V_{\mathscr{M}}(X_t) - C_t - R_t - \lambda I_t\right)\right], \tag{1.1}$$

sous les contraintes

$$X_t \in K_{\mathscr{P}_t} \tag{1.2}$$
$$\mathscr{A}_t \subseteq \mathscr{Q}(\mathscr{E}_t) \cap \mathscr{M}, \tag{1.3}$$
$$\mathscr{A}_{t^+} = \emptyset \quad \text{à la succession}, \tag{1.4}$$
$$\mathsf{Adm}(\Theta_t) = 1 \Rightarrow \mathsf{PreuveFraiche}(\Theta_t) = 1, \tag{1.5}$$
$$\mathsf{RetourArriere}(X_t) \text{ demeure exécutable.} \tag{1.6}$$

Il ne suffit pas à l'institution de devenir capable ; elle doit demeurer gouvernable pendant qu'elle change.

2 Périmètre institutionnel et exigences

2.1 Quatre familles d'exigences

Une architecture de Civilisation-machine exige une identité persistante, des preuves typées, une autorité explicite et une succession vérifiable. La racine doit publier sa version, ses rôles et ses limites. Toute affirmation matérielle doit avoir une source, un dénominateur complet, une date, un sujet versionné et une voie de contestation. Toute autorité doit préciser le sujet, l'action, la cible, les plafonds, la durée, la surveillance, la révocation et le retour arrière. Toute succession doit identifier ce qui est hérité, ce qui change et la preuve nouvelle.

2.2 Exigences non fonctionnelles

L'institution doit aussi être utilisable : lisibilité humaine, données structurées, portabilité, activation minimale des rôles, progression A0–A4, échec fermé, reproductibilité des versions et accès bilingue. Une architecture formelle incompréhensible sera contournée dans l'exploitation quotidienne.

2.3 Rapport aux cadres existants

Le cadre NIST AI RMF organise la gestion du risque autour de Govern, Map, Measure et Manage, et son profil pour l'IA générative adapte ces fonctions aux risques propres à ces systèmes [7, 61]. Les initiatives NIST de 2026 sur les agents insistent sur l'identité, l'autorisation, l'interopérabilité et la sécurité [11, 51]. L'architecture zéro confiance refuse la confiance ambiante et exige une autorisation continue et spécifique aux ressources [14, 58]. SUCCESSOR Ω ajoute une question de succession : pourquoi une implémentation remplaçante aurait-elle le droit d'hériter des conditions de l'autorisation antérieure ?

La proposition croise aussi les systèmes multi-agents, la conception des mécanismes, l'IA coopérative et la gouvernance des ressources communes [18, 50, 54, 60, 66]. Sa contribution distinctive est la composition explicite des transitions entre identité, travail, règlement, droits, acceptation, preuve, admission, autorité, Chronique et succession.

3 Une architecture, quatre racines irréductibles

3.1 La séparation sémantique comme levier institutionnel

Chaque identité reçoit une fonction qu'aucune autre ne remplace. QUEBEC.AI répond à la question de la finalité. MONTREAL.AI répond à celle de la constitution et de l'exploitation de l'institution. AGI.eth répond à celle des agents capables d'agir, de concourir, de vérifier et de régler. ASI.eth répond à celle de la continuité constitutionnelle, de la preuve, de l'autorité et de la succession.

Table 3.1 : Architecture institutionnelle à quatre racines.

Racine	Rôle canonique	Question principale	Fonction de plus grande valeur
QUEBEC.AI	Mission et voix publique	Que doit servir l'intelligence ?	Finalité, légitimité bilingue, éducation, concertation et origine des missions
MONTREAL.AI	Constructeur et administrateur	Qui constitue et exploite l'institution ?	GoalOS, architecture de preuve, déploiement client, gouvernance et assurance
AGI.eth	Racine des agents et de l'économie	Quels agents peuvent agir, concourir et régler ?	Identité, emplois, validateurs, nœuds, séquestre, AGIALPHA et reçus
ASI.eth	Racine constitutionnelle	Qu'est-ce qui persiste et qui peut agir ?	Pacte, successeur, preuve, admission, autorité, Chronique et retour arrière

Figure 3.1 : Les quatre racines constituent une seule pile allant de la mission à la civilisation. Leurs pouvoirs sont complémentaires et doivent demeurer distincts.

Cette séparation empêche l'effondrement des catégories. La mission publique n'est pas l'implémentation commerciale ; l'exécution économique n'est pas l'autorité constitutionnelle. Elle permet aussi une confiance différenciée : un client peut utiliser un déploiement MONTREAL.AI sans conférer à ASI.eth un pouvoir opérationnel, ou recourir à AGI Jobs sans adopter toutes les positions publiques de QUEBEC.AI.

3.2 QUEBEC.AI : la couche de mission

QUEBEC.AI exprime ce que l'intelligence avancée devrait préserver et accroître : liberté, prospérité, culture, connaissance, autonomie institutionnelle, coopération et biens communs. Sa fonction optimale est normative et rassembleuse plutôt qu'opérationnelle. Elle peut publier une doctrine bilingue, rendre les travaux de pointe accessibles, convoquer des forums et faire émerger des missions dont les résultats seront ensuite soumis à la preuve.

Une constitution de mission minimale est

$$\mathscr{M} = \langle \text{Finalite}, \text{Beneficiaires}, \text{Objectifs}, \text{NonObjectifs}, \text{Contraintes}, \text{Mesures}, \text{Appel}, \text{Sortie} \rangle. \tag{3.1}$$

Les non-objectifs sont essentiels : un optimiseur peut améliorer une métrique en détruisant ce que l'institution devait protéger. L'appel et la sortie distinguent la coordination du contrôle.

3.3 MONTREAL.AI : la couche institutionnelle

MONTREAL.AI transforme une mission en objet gouverné au moyen de GoalOS, Proof Gradient, AGI Jobs, SUCCESSOR Ω et des protocoles MCAP. Le produit durable n'est pas un modèle particulier, mais une intelligence de mission appartenant au client : constitution, programmes du monde, preuves, tests, politiques, mémoire admise, Chronique et contrôles de déploiement.

L'institution fondatrice peut conserver la propriété et la continuité du parent tout en séparant la vérification finale, l'acceptation du client et l'admission à conséquences. La force de l'architecte ne doit pas devenir l'auto-certification de toutes ses affirmations.

3.4 AGI.eth : la couche d'action et d'économie

AGI.eth rend adressables les agents, emplois, validateurs, outils, nœuds et surfaces de règlement. Une identité sémantique peut résoudre vers un contrôleur, des capacités, des points d'accès, une version, des preuves, des limites et un état de révocation.

L'économie ne se réduit pas au paiement. Le séquestre rend l'enjeu explicite ; le cautionnement expose un participant à une conséquence définie ; la récompense rémunère l'évaluation ; le règlement fixe une conséquence économique ; le NFT d'achèvement peut servir de reçu. Aucun de ces éléments ne prouve à lui seul les droits, la qualité hors chaîne ou l'indépendance.

Le cycle est

$$\begin{aligned} &\text{Mission} \to \text{Emploi} \to \text{Sequestre} \to \text{Execution} \to \text{Preuve} \\ &\qquad \to \text{Validation} \to \text{Acceptation/Litige} \to \text{Reglement} \to \text{Chronique}. \end{aligned} \tag{3.2}$$

Exécuter, valider, accepter, admettre et régler sont des actes différents.

3.5 ASI.eth : la couche civilisationnelle adressable

ASI.eth n'est pas prétendu être l'ASI. C'est la racine proposée d'une architecture de Civilisation-machine. Un client indépendant devrait y découvrir le Manifeste racine, le Pacte SUCCESSOR Ω, le registre des rôles, le successeur courant, les preuves, les vérificateurs, les décisions d'admission, les enveloppes d'autorité, les événements de Chronique et les limites.

Les sous-noms canoniques peuvent inclure successor.asi.eth, covenant.asi.eth, registry.asi.eth, chronicle.asi.eth, proof.asi.eth, verifier.asi.eth, admission.asi.eth, policy.asi.eth, rollback.asi.eth et treasury.asi.eth. Ils ne devraient être activés publiquement que lorsque leur fonction est réelle et leur contrôle divulgué.

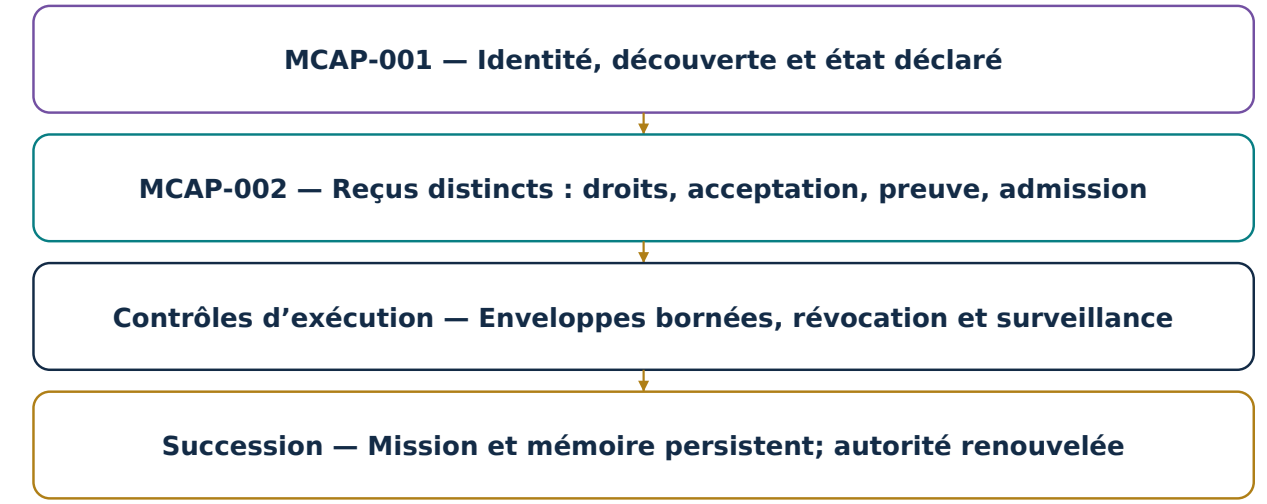


Figure 3.2 : La pile protocolaire sépare adressabilité, travail économique, admission probante, autorité et succession.

3.6 La boucle institutionnelle

La légitimité de mission produit une demande qualifiée ; la demande produit des déploiements gouvernés ; les déploiements produisent des emplois, des preuves et des règlements ; la preuve admise produit une meilleure mémoire ; la mémoire améliore les successeurs ; les successeurs créent davantage de valeur de mission. Chaque transition garde cependant son frein probant. Sans lui, la mémoire amplifie l'erreur et la marque remplace la preuve.

Principe de conception 3.1 (Rareté sémantique)**.** Utiliser peu d'identités canoniques et leur donner des fonctions précises. Réserver d'autres noms au besoin, mais n'activer que ceux dont le rôle, le contrôleur, le statut et les preuves sont inspectables.

3.7 Propriété sans domination

Le client peut posséder et exporter sa mission, sa mémoire admise, ses références probantes, ses politiques, sa Chronique et son état de succession sans acquérir les logiciels génériques, les marques ou les racines de MONTREAL.AI. Inversement, MONTREAL.AI peut préserver l'architecture tout en ne disposant d'aucune autorité unilatérale sur les opérations conséquentes d'un client. La portabilité est un droit conçu ; l'autorité demeure spécifique à la mission.

4 État institutionnel et opérateur de succession

PARTIE II

THÉORIE FORMELLE DE LA SUCCESSION

4.1 Un état institutionnel typé

L'état complet d'une institution de mission au temps t est

$$X_t = \big(\mathscr{M}_t, \mathscr{K}_t, \mathscr{E}_t, \mathscr{P}_t, \mathscr{A}_t, \mathscr{C}_t, \mathscr{R}_t, \Theta_t\big). \tag{4.1}$$

Les composantes sont volontairement typées : $\mathscr{M}$ contient la mission et les obligations ; $\mathscr{K}$ la mémoire admise ; $\mathscr{E}$ les graphes et reçus de preuve ; $\mathscr{P}$les politiques, droits et contraintes ; $\mathscr{A}$ l'autorité courante ; $\mathscr{C}$ la Chronique ; $\mathscr{R}$ les rôles et contrôleurs ; Θ les modèles, outils, invites, adaptateurs et paramètres d'exécution.

Le typage empêche les promotions silencieuses. Une sortie de modèle ne devient pas connaissance parce qu'elle est copiée dans une base. Un résultat probant ne devient pas autorité parce qu'un seuil numérique est dépassé. Une transaction ne devient pas acceptation du client. Chaque changement de type exige une transition, un acteur, un paquet et une trace.

4.2 L'opérateur de succession

Soit

$$\mathscr{I}_t = \big(\mathscr{M}_t, \mathscr{K}_t^{\mathrm{adm}}, \mathscr{O}_t, \mathrm{Refs}(\mathscr{C}_t)\big) \tag{4.2}$$

l'ensemble héritable, où $\mathscr{O}_t$ représente les obligations survivantes. L'état du successeur avant admission est

$$\mathsf{S}(X_t, \Theta_{t+1}) = \big(\mathscr{M}_t, \mathscr{K}_t^{\mathrm{adm}}, \mathscr{E}_t^{\mathrm{hist}}, \mathscr{P}_{t+1}, \emptyset, \mathscr{C}_t \,\|\, c_{\mathrm{nom}}, \mathscr{R}_{t+1}, \Theta_{t+1}\big). \tag{4.3}$$

La preuve historique peut accompagner la mémoire, mais elle n'est pas la preuve courante de la nouvelle implémentation. L'autorité est réinitialisée à l'ensemble vide.

Proposition 4.1 (Non-héritage de l'autorité)**.** *Supposons que toute action conséquente exige une enveloppe signée par un Principal autorisé et liée au condensat et à la génération de l'implémentation, avec nouvelle preuve fraîche requise pour chaque admission. Si la succession change le condensat et pose $\mathscr{A}_{t+1} = \emptyset$, le successeur ne peut exercer valablement l'autorité du prédécesseur sans nouvelle admission.*

Démonstration. L'enveloppe antérieure lie le condensat d_t, tandis que le successeur présente $d_{t+1} \neq d_t$. La vérification du sujet échoue. Comme aucune enveloppe courante n'existe, une nouvelle preuve et une nouvelle admission sont nécessaires. La conclusion dépend d'une passerelle d'action complète, sans voie de contournement. □

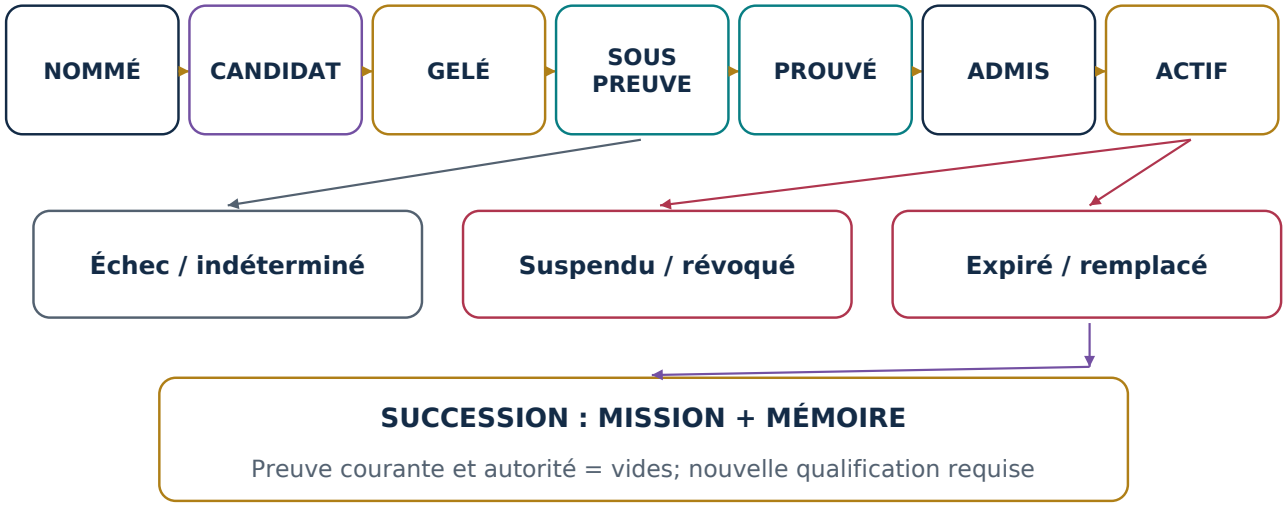


FIGURE 4.1 : Machine à états de succession. La mission et la mémoire admise peuvent franchir la frontière générationnelle ; l'autorité courante ne le peut pas.

4.3 Admission de la mémoire

Un élément m n'entre dans la mémoire institutionnelle que si

$$\mathsf{AdmettreMemoire}(m) = \mathsf{Integrite}(m) \wedge \mathsf{Droits}(m) \wedge \mathsf{Preuve}(m) \wedge \mathsf{Portee}(m) \wedge \mathsf{Conservation}(m). \tag{4.4}$$

Un score souple de conservation peut combiner poids probant w_i, pertinence r_i, qualité q_i et ancienneté s_i :

$$\rho_i = w_i r_i q_i e^{-\lambda s_i}, \tag{4.5}$$

mais aucun score ne compense l'absence d'un droit matériel ou une corruption d'intégrité. Les éléments rejetés peuvent être mis en quarantaine afin de conserver leur histoire sans leur attribuer le statut de connaissance.

4.4 Identité versionnée et lignée

Une identité persistante s'appuie sur une racine n et un graphe de lignée

$$G_{\text{lin}} = (V_\Theta, E_{\text{succ}}), \tag{4.6}$$

où les sommets sont des engagements immuables d'implémentation et les arêtes des événements de succession. Le nom lisible peut résoudre vers le sommet courant, tandis que la Chronique préserve les sommets antérieurs. Ainsi, la mutabilité du nom n'efface pas l'histoire.

4.5 Équivalence institutionnelle

Deux implémentations sont équivalentes seulement relativement à une mission, une distribution de tests $\mathscr{T}$, un ensemble d'actions $\mathscr{U}$ et une tolérance ε :

$$\Theta_i \sim_{\mathscr{M},\mathscr{T},\mathscr{U},\varepsilon} \Theta_j. \tag{4.7}$$

Cette relation exige des distributions de résultats proches, des violations de politique comparables et aucune autorité supplémentaire. Elle inclut les risques de queue et les modes d'échec, non seulement la performance moyenne.

4.6 Avantage de mission

Un modèle plus récent ne mérite pas automatiquement l'admission. Définissons

$$\alpha(\Theta_c) = J(\Theta_c) - \max\{J(\Theta_{\text{courant}}), J(\Theta_\beta)\} - C_{\text{transition}} - R_{\text{queue}} - F_{\text{fragilite}}. \tag{4.8}$$

Un α positif est nécessaire, mais insuffisant. Les droits, la preuve, l'autorité et le retour arrière restent des portes indépendantes.

Principe de conception 4.2 (Aucune prime à la nouveauté)**.** La récence, le nombre de paramètres, le prestige d'une marque ou la réputation sur des bancs d'essai généraux ne prouvent pas l'avantage de mission. Le candidat doit surpasser la meilleure solution disponible après coûts, risques, fragilité et requalification.

5 Preuve, admission et autorité

5.1 La preuve comme graphe typé

Le graphe probant

$$G_{\mathscr{E}} = (V_{\mathscr{E}}, E_{\text{sup}}, E_{\text{conf}}, E_{\text{dep}}) \tag{5.1}$$

relie affirmations, artefacts, tests, observations, signatures, droits et limites. Les arêtes de dépendance révèlent les sources communes. Dix rapports produits à partir de la même base cachée ne sont pas dix preuves indépendantes.

Pour des poids w_i et une matrice de dépendance Σ, une mesure prudente est

$$W_{\text{eff}} = \frac{(1^\top \mathrm{w})^2}{\mathrm{w}^\top \Sigma \mathrm{w} + \epsilon}. \tag{5.2}$$

Elle ne constitue pas une norme universelle ; elle oblige à représenter la corrélation plutôt qu'à compter des documents.

5.2 Classes de preuve

Table 5.1 : Classes de preuve MCAP-002.

Classe	Nom	Signification	Exigence de promotion
E0	Déclarée	Affirmation attribuable sans corroboration	Source, version et artefact inspectable
E1	Fondée sur une source	Source attribuable, non rejouée indépendamment	Reproduction indépendante
E2	Reproductible	Résultat reproductible par un tiers	Contrôle indépendant et conflits divulgués
E3	Contrôle indépendant	Vérificateur, client ou gardien hors contrôle du demandeur	Conséquence mesurable, droits libérés, autorité bornée
E4	À conséquence	Résultat accepté et observable sous gouvernance	Répétition, surveillance et succession réussie

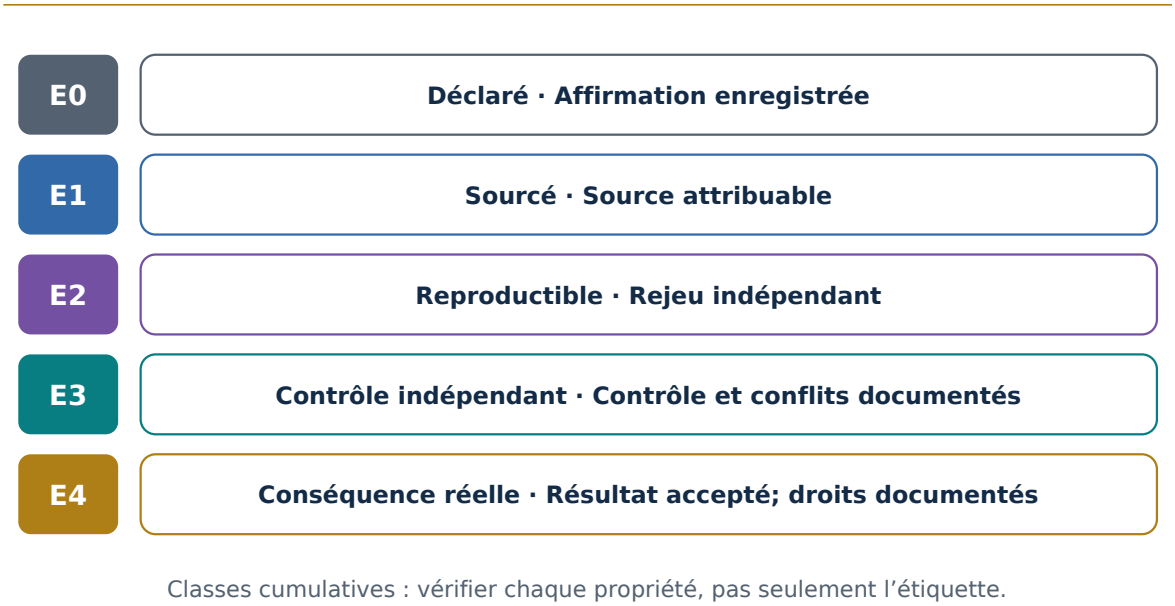


Figure 5.1 : L'échelle probante exige à chaque transition une propriété absente de la classe précédente.

Un condensat, un NFT, un nom ENS, un jeton ou un horodatage peut renforcer l'intégrité sans changer automatiquement la classe. Une affirmation E0 inscrite sur chaîne reste E0, bien qu'elle soit solidement horodatée.

5.3 Le paquet de preuve

Un paquet de preuve lie

$$\mathsf{P} = \langle \mathscr{M}, d_\Theta, T, D, M, B, E, L, V, S, t_0, t_1 \rangle, \tag{5.3}$$

avec candidat figé, plan d'essai, dénominateur complet, mesures, références, éléments probants, limites, identité et conflits du vérificateur, signatures et intervalle de validité.

Règle de gel avant preuve

Le plan final et le condensat du candidat doivent être engagés avant l'évaluation décisive. Toute modification matérielle crée une nouvelle version ou demeure explicitement exploratoire.

Le dénominateur complet inclut succès, échecs, abstentions, exclusions et essais invalides. Une preuve qui ne publie que les cas favorables est structurellement biaisée.

5.4 Interprétation bayésienne et valeur de la preuve

Pour une hypothèse H_1 et son alternative H_0,

$$\log \frac{P(H_1 \mid e)}{P(H_0 \mid e)} = \log \frac{P(H_1)}{P(H_0)} + \log \frac{P(e \mid H_1)}{P(e \mid H_0)}. \tag{5.4}$$

La reproduction indépendante apporte de l'information lorsque les sources ne sont pas corrélées. La valeur décisionnelle de la preuve est

$$\mathrm{VoP}(\mathsf{P}) = \mathbb{E}[U(d^*(\mathsf{P}), \theta)] - \mathbb{E}[U(d_0, \theta)] - C(\mathsf{P}). \tag{5.5}$$

La meilleure preuve n'est pas nécessairement la plus volumineuse, mais celle qui modifie une décision importante à un coût acceptable.

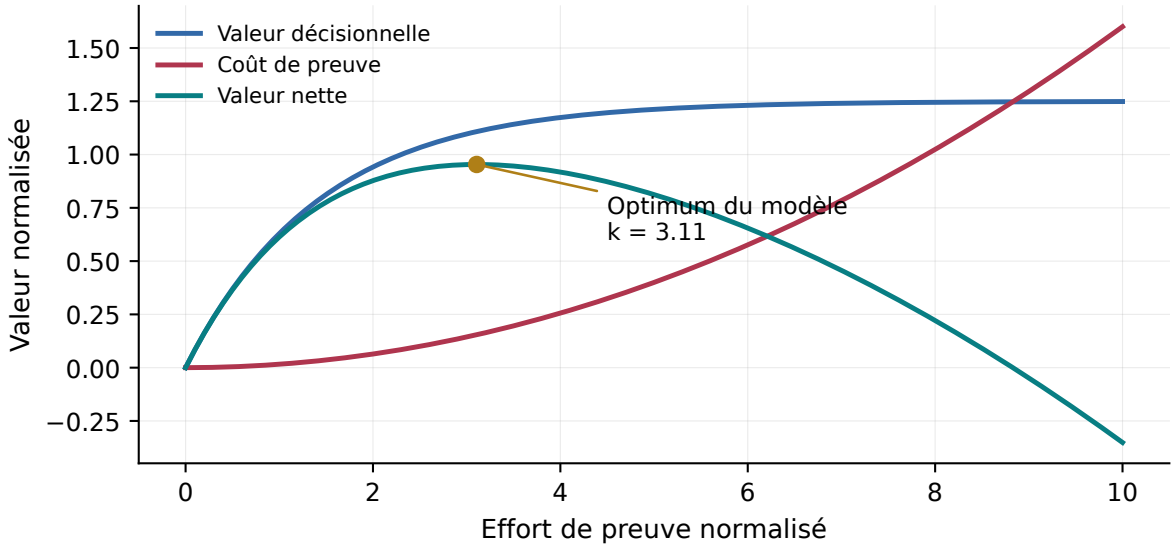


Figure 5.2 : Fonctions déclarées : bénéfice $1.25(1 - e^{-0.7k})$, coût $0.016k^2$ et différence. L'optimum est propre à ce modèle synthétique, non une estimation de valeur client.

5.5 Admission et enveloppe d'autorité

La preuve décrit ce que soutiennent les faits. L'admission décrit ce que l'institution décide d'autoriser. Une enveloppe est

$$\mathscr{A} = \langle s, m, U, T, B, \tau_0, \tau_1, O, R, K, \mathsf{P}, \mathsf{Adm}, \sigma \rangle, \tag{5.6}$$

avec sujet, mission, actions, cibles, limites, durée, observabilité, révocation, retour arrière, références probantes et signature du Principal.

L'ordre partiel $\preceq$ représente l'inclusion de tous les champs. L'autorité admise doit satisfaire

$$\mathscr{A} \preceq \mathcal{Q}(\mathsf{P}), \qquad \mathscr{A} \preceq \mathscr{M}, \qquad \mathscr{A} \preceq \mathscr{P}. \tag{5.7}$$

Davantage de preuve peut justifier une autorité plus large, mais jamais au-delà de la mission ou de la politique.

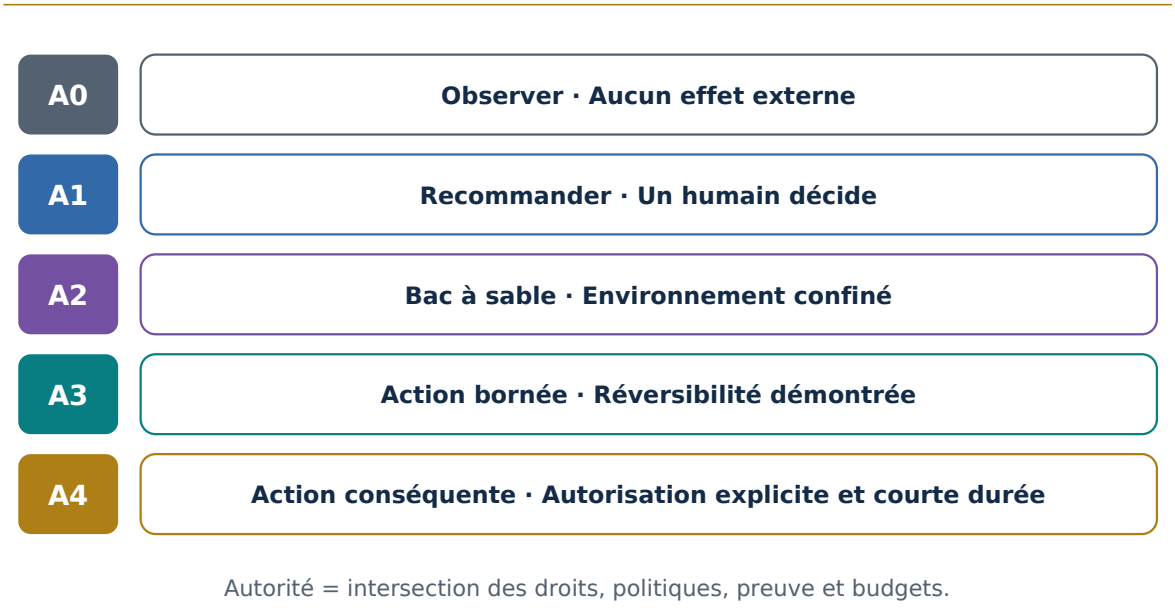


Figure 5.3 : Échelle d'autorité par défaut. La succession retourne le nouveau sujet à A0 jusqu'à sa requalification.

5.6 Échelle A0–A4

- A0 Observer, résoudre, simuler et recommander sans effet opérationnel.
- A1 Émettre une recommandation attribuable à une personne responsable.
- A2 Agir dans un bac à sable sans effet externe non contrôlé.
- A3 Exécuter une action bornée et réversible en production.
- A4 Exécuter une action conséquente explicitement approuvée, étroite et de courte durée.

La référence v9 fournit la formation A0 et la répétition A1/A2. Les effets externes conséquents sont désactivés.

Précision : l'autorité est un ensemble, pas un score

Les droits, la politique, le risque et la preuve définissent des ensembles d'actions contextualisées. L'enveloppe admise doit être un sous-ensemble de leur intersection ; elle n'est pas un produit de quantités hétérogènes. Une requête doit satisfaire une enveloppe valide complète. La cible d'une autorisation ne peut être combinée avec le budget ou l'action d'une autre. La génération doit être liée au sujet même si le code demeure identique. Les pénalités de transition et de risque ne doivent pas compter deux fois un coût déjà retranché de la valeur de mission.

6 Chronique, viabilité et retour arrière

6.1 La mémoire institutionnelle des conséquences

Une base de données conserve un état. Une Chronique conserve la raison pour laquelle un état conséquent a été autorisé à changer. Elle forme une séquence append-only d'engagements portant sur les admissions, actions, échecs, réparations, suspensions, révocations et successions. Les preuves sensibles peuvent rester protégées, mais leur existence, leur gardien, leur condensat, leur statut et leur emploi dans une décision doivent être attribuables.

Pour un événement canonique m_i, un horodatage t_i, des signataires S_i, des références E_i et un état d'autorité A_i, la chaîne est

$$h_i = H(\text{domaine} \parallel h_{i-1} \parallel \text{canon}(m_i, t_i, S_i, E_i, A_i)) . \tag{6.1}$$

Les événements peuvent être regroupés dans une racine de Merkle et horodatés auprès de gardiens indépendants [24, 40].

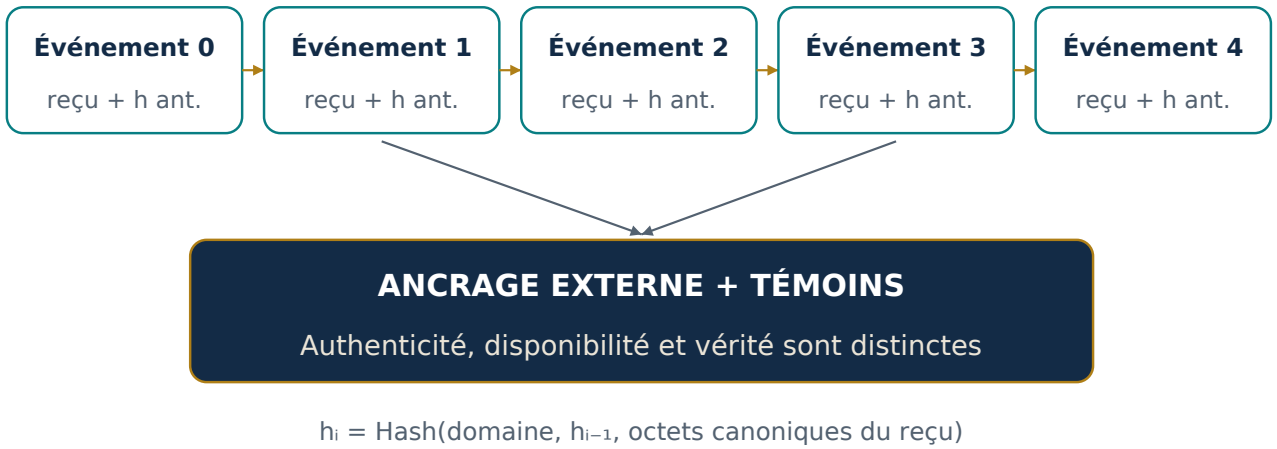


Figure 6.1 : La Chronique relie les transitions institutionnelles. L'ancrage externe et la garde séparée renforcent la résistance à la réécriture rétrospective.

Proposition 6.1 (Détection de modification). *Si la sérialisation est déterministe et la fonction H résistante aux collisions, la modification d'un événement accepté change son condensat et tous les condensats ultérieurs. Si un*

gardien indépendant conserve un condensat ultérieur, la modification rétrospective non détectée est calculatoirement impraticable selon l'hypothèse cryptographique.

Démonstration. Le message canonique modifié change l'entrée de H, sauf collision. Le nouveau h_j modifie ensuite h_{j+1} par induction. Un condensat terminal conservé indépendamment révèle la divergence. La propriété démontre l'intégrité relative de l'histoire, non la vérité de l'événement initial. □

6.2 Taxonomie des événements

Les événements utiles incluent la constitution de mission, l'importation d'un règlement, la revue des droits, l'acceptation du client, le gel du candidat, le début et le résultat de la preuve, l'admission, l'émission et l'expiration de l'autorité, le retour arrière, l'admission ou la quarantaine de mémoire, la nomination et la qualification d'un successeur. Chaque événement référence les dépendances antérieures sans les confondre.

6.3 Viabilité sous autorité bornée

Considérons

$$x_{t+1} = f(x_t, u_t, w_t), \tag{6.2}$$

où u_t est une action autorisée et w_t une perturbation. L'ensemble sûr est

$$K = \{x \,:\, g_j(x) \le 0,\ j = 1, \dots, m\}. \tag{6.3}$$

Le noyau de viabilité regroupe les états depuis lesquels une politique admissible peut maintenir le système dans K pour toute perturbation modélisée [5]. Une fonction barrière B soutient la condition

$$B(x) \le 0 \Rightarrow \exists u \in U_{\mathscr{A}}(x) \,:\, B(f(x,u,w)) \le 0 \quad \forall w \in W. \tag{6.4}$$

Si la passerelle ne peut établir cette propriété, elle doit s'abstenir ou demander une action plus étroite.

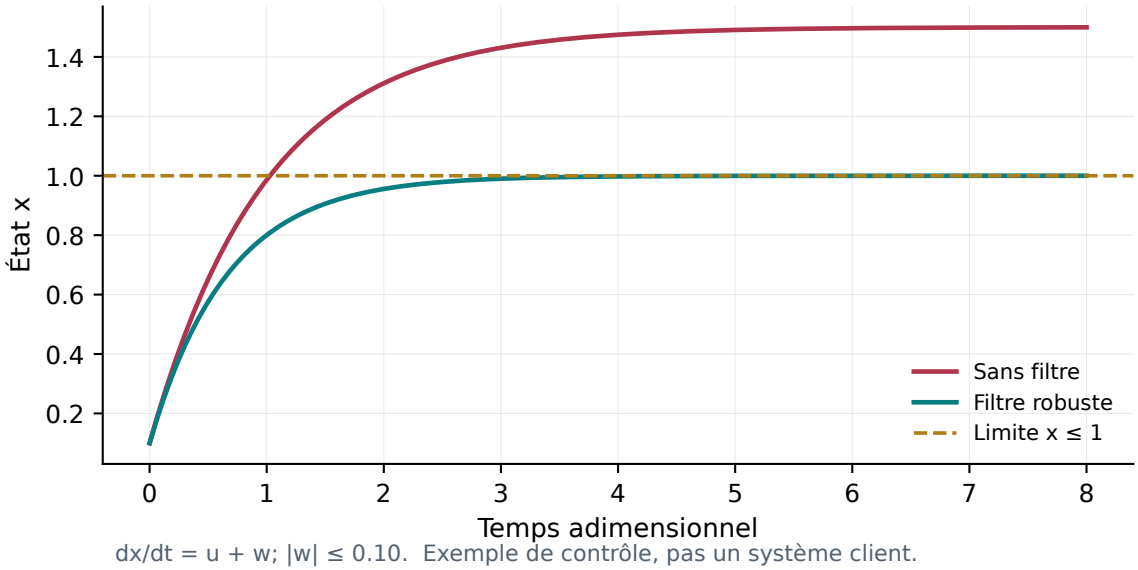


Figure 6.2 : Exemple explicite $\dot{x} = u + w$, $|w| \leq 0.10$, frontière $x \leq 1$. Le filtre impose $u \leq 1.5(1 - x) - 0.10$ lorsque ce contrôle est faisable. Ce nést pas un noyau de viabilité calculé pour un système client.

6.4 Stabilité et retour arrière

Une fonction de Lyapunov $V(x)$ peut soutenir un argument local lorsque

$$V(x) > 0 \ (x \neq x^*), \qquad \Delta V(x) = V(f(x, \pi(x), w)) - V(x) < 0 \tag{6.5}$$

dans une région et pour un ensemble de perturbations déclarés [31, 38]. En pratique, V peut agréger écart de politique, exceptions non résolues, preuve périmée, actions non revues et distance d'un état récupérable.

Un plan de retour arrière n'est crédible qu'après exercice. Le reçu doit indiquer le déclencheur, l'état de l'autorité, les étapes, le temps écoulé, les effets résiduels, les pertes de preuve et les mesures correctives. Restaurer une base de données n'efface pas nécessairement une communication publique, un paiement ou un dommage relationnel.

6.5 Budget d'irréversibilité

Pour un coût d'irréversibilité $I(u, x)$, l'enveloppe impose

$$I(u, x) \leq I_{\max}(\mathscr{A}). \tag{6.6}$$

Le plafond ne monte qu'avec une preuve plus forte et une approbation responsable. Même une recommandation A1 peut avoir une forte influence sociale ; l'interface et les défauts de présentation font donc partie du système de contrôle.

Le retour arrière n'efface pas les conséquences

Un retour arrière peut restaurer l'état technique tout en laissant des conséquences économiques, juridiques, réputationnelles ou humaines. Le reçu doit distinguer état restauré et conséquence résiduelle.

7 Conception des mécanismes et stabilité stratégique

7.1 L'institution comme mécanisme répété

Une Civilisation-machine est un mécanisme répété entre Principaux, clients, agents, vérificateurs, opérateurs, gardiens et fournisseurs de ressources. Chaque participant possède des informations privées, des intérêts propres et des conflits possibles. La gouvernance ne peut supposer la bienveillance ; elle doit rendre l'action utile et véridique stratégiquement préférable lorsque cela est possible.

Pour des types θ_i, rapports r_i, allocation $g(r)$ et transfert $p_i(r)$,

$$u_i(\theta_i, r) = v_i(g(r), \theta_i) - p_i(r) - \ell_i(r), \tag{7.1}$$

où ℓ_i comprend perte de caution, réputation, occasion et conséquence contractuelle. La théorie de la conception des mécanismes étudie les règles capables de produire des résultats souhaités malgré le comportement stratégique [28, 50, 64].

7.2 Propriétés institutionnelles

Les propriétés recherchées comprennent rationalité individuelle, compatibilité incitative, solvabilité, absence d'auto-certification unilatérale, résistance aux coalitions, appel, sortie et autorité bornée. Aucune conception ne satisfait toutes les propriétés pour toute structure d'information. L'obligation est de publier les compromis et de savoir quels échecs demeurent rentables.

7.3 Jeu de séparation des pouvoirs

Soient un demandeur C, un vérificateur V et un Principal P. Le demandeur reçoit un bénéfice b de l'admission et peut engager un coût m pour manipuler. Le vérificateur reçoit une rémunération r_v, supporte un coût de diligence d et peut recevoir un pot-de-vin β. La séparation réduit la probabilité de succès de la manipulation de q_c à q_s et augmente la sanction attendue s.

La manipulation devient défavorable lorsque

$$q_s b - m - (1 - q_s)s < b - c_{\text{honnete}}, \tag{7.2}$$

et la collusion du vérificateur lorsque

$$\beta - d_{\text{collusion}} - p_{\text{detection}} s_v < r_v - d. \tag{7.3}$$

Des portefeuilles distincts ne changent ces inégalités que si le contrôle, l'accès à la preuve, les incitations et la contestation sont effectivement séparés.

Des clés distinctes ne prouvent pas des contrôleurs indépendants.

FIGURE 7.1 : La séparation des rôles change les gains stratégiques seulement lorsque contrôle, preuve, incitations et contestation sont réellement distincts.

7.4 Équilibre corrélé

Une distribution $\pi(a)$ constitue un équilibre corrélé lorsque, pour tout joueur i et toute déviation a_i',

$$\sum_{a_{-i}} \pi(a_i, a_{-i})[u_i(a_i, a_{-i}) - u_i(a_i', a_{-i})] \geq 0. \tag{7.4}$$

Un plan de preuve public peut jouer le rôle de dispositif de corrélation : il coordonne les objets évalués et les règles de décision sans imposer un nœud omniscient [6]. La stabilité stratégique ne garantit toutefois pas la vérité ; un équilibre de collusion demeure possible.

7.5 Coopération répétée

Dans un dilemme du prisonnier répété, avec récompense R, tentation T, punition P et facteur d'actualisation δ, une coopération de type déclencheur est soutenable si

$$\delta \geq \frac{T - R}{T - P}. \tag{7.5}$$

L'identité persistante augmente l'horizon effectif ; la preuve indépendante augmente la détection ; la perte d'accès réduit l'attrait de la défection. L'objectif n'est pas une sévérité arbitraire, mais la crédibilité de l'interaction répétée [8].

7.6 Dynamiques évolutives

Pour des parts de stratégies x_i et une matrice A,

$$\dot{x}_i = x_i[(Ax)_i - x^\top Ax]. \tag{7.6}$$

La stratégie coopérative croît si son gain dépasse la moyenne [39, 62]. Les récompenses, la réputation, le séquestre et l'accès modifient la matrice ; ils ne garantissent pas la coopération.

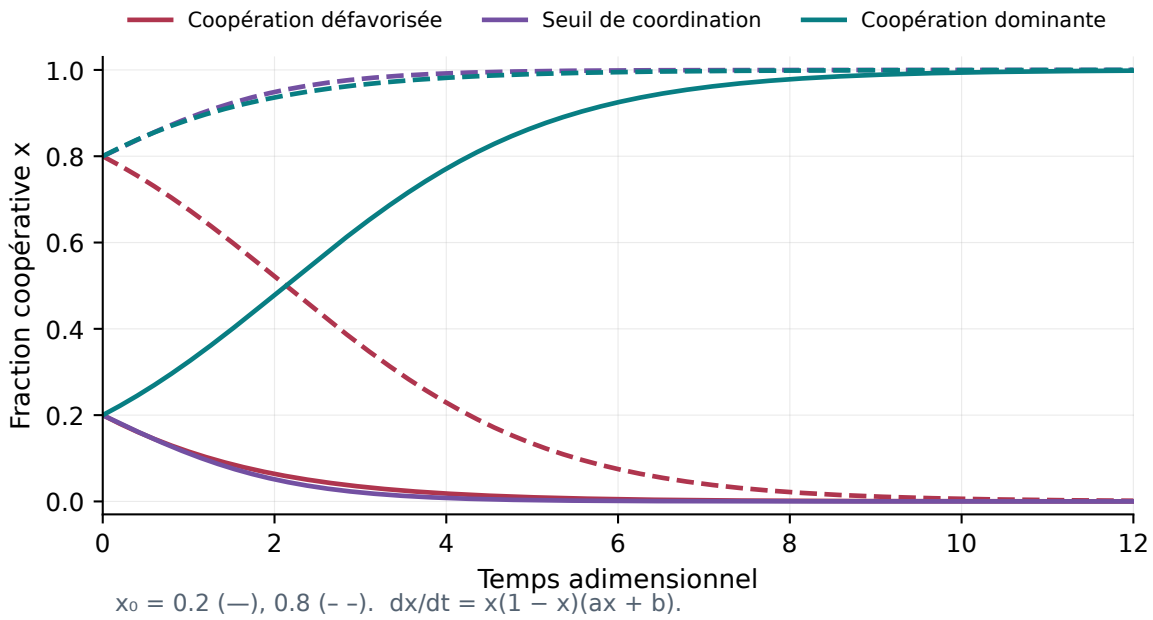


Figure 7.2 : Modèle à deux stratégies : $\dot{x} = x(1-x)(ax+b)$. Les trois régimes de gains sont déclarés dans le code ; les trajectoires partent de 0.2 et 0.8. Ces courbes synthétiques ne mesurent pas une adoption réelle.

Principe de conception 7.1 (Rendre la prédation stratégiquement peu profonde)**.** L'institution ne suppose pas que les stratégies prédatrices disparaissent. Elle les rend observables, coûteuses, pauvres en accès et incapables d'acquérir une autorité persistante. Les stratégies coopératives accèdent durablement à la connaissance, à la réputation et aux missions futures par contribution probante.

8 Physique statistique de la stabilité institutionnelle

8.1 Trois niveaux de prétention

La physique statistique fournit des outils pour les distributions, les fluctuations, l'ordre collectif et la métastabilité. Elle n'établit pas qu'une institution sociale est un matériau thermodynamique. Ce chapitre distingue une identité exacte sur un ensemble fini, un processus stochastique hypothétique et une interprétation institutionnelle. Les prédictions comportementales exigent une loi de transition et un étalonnage empirique.

8.2 Ensembles institutionnels admissibles

Un état fini s représente agents, preuves, rôles, obligations et permissions. Définissons un coût dans une unité déclarée de perte de mission :

$$H(s) = C(s) + \lambda_R R(s) + \lambda_I I(s) + \lambda_C C_{\text{capture}}(s) - \lambda_V V_{\mathscr{M}}(s) - \lambda_Q Q_{\mathscr{E}}(s). \tag{8.1}$$

Les coefficients convertissent les quantités hétérogènes dans cette unité. Un faible coût modélisé n'est pas une légitimité morale. Les interdictions définissent séparément un support non vide $\mathcal{S}_{\text{adm}}$. Une pénalité finie, même élevée, n'impose pas une interdiction : à température effective positive, chaque état de coût fini du support conserve une probabilité positive.

Pour $\tau > 0$ dans la même unité que H, posons $\beta = \tau^{-1}$:

$$Z_\tau = \sum_{s \in \mathcal{S}_{\text{adm}}} e^{-H(s)/\tau}, \qquad p_\tau(s) = Z_\tau^{-1} e^{-H(s)/\tau}, \qquad F_\tau = -\tau \log Z_\tau. \tag{8.2}$$

L'exposant est sans dimension. τ représente un bruit de sélection, non une température mesurée en kelvins. L'analogie s'inspire de l'ensemble canonique [21].

8.3 Principe variationnel de Gibbs sur un ensemble fini

Pour toute distribution q sur le support admissible,

$$\mathscr{F}_\tau[q] = \sum_s q(s) H(s) + \tau \sum_s q(s) \log q(s), \qquad 0 \log 0 = 0. \tag{8.3}$$

Proposition 8.1 (Identité variationnelle finie)**.** *Si le support est fini et non vide, les coûts y sont finis et $\tau > 0$, alors*

$$\mathscr{F}_\tau[q] - F_\tau = \tau D_{\text{KL}}(q \| p_\tau) \geq 0. \tag{8.4}$$

L'unique minimum est atteint pour $q = p_\tau$.

Démonstration. Insérer $\log p_\tau(s) = -H(s)/\tau - \log Z_\tau$ dans la divergence et utiliser $\sum_s q(s) = 1$. La positivité de l'entropie relative donne l'inégalité ; l'égalité exige des distributions identiques. Cette identité algébrique ne prouve pas que le comportement réel suit p_τ. □

8.4 Une loi de transition reste nécessaire

Sur un graphe fini admissible connexe, un noyau de proposition symétrique et la règle de Metropolis

$$A(s, s') = \min\{1, e^{-[H(s') - H(s)]/\tau}\} \tag{8.5}$$

satisfont l'équilibre détaillé. L'irréductibilité et une boucle d'attente apériodique permettent la convergence. Des classes déconnectées, des cycles déterministes ou des incitations dirigées peuvent empêcher cette conclusion. Une institution stratégique générale peut être hors équilibre.

Un jeu à potentiel exact admet une construction apparentée pour des mises à jour logit asynchrones, à joueur unique et température commune. Les différences unilatérales d'utilité doivent réellement correspondre aux différences du potentiel [42] ; un score de tableau de bord ne suffit pas.

8.5 Entropie, diversité et incertitude

L'entropie de Shannon institutionnelle est sans dimension [59]. Si R représente le contrôle, E l'état probant et A l'action,

$$\mathsf{H}(R, E, A) = \mathsf{H}(R) + \mathsf{H}(E \mid R) + \mathsf{H}(A \mid E, R). \tag{8.6}$$

Cette règle ne constitue pas une somme arbitraire de scores de diversité, d'incertitude et d'opacité. Les dépendances importent. Une diversité de vérificateurs peut être souhaitable alors que l'incertitude sur leur contrôle bénéficiaire ne l'est pas. Minimiser l'entropie totale n'est donc pas un objectif institutionnel suffisant.

8.6 Paysages de Landau et métastabilité

Pour un paramètre d'ordre signé m, étudions

$$F(m; a, h) = \frac{1}{4}m^4 + \frac{1}{2}am^2 - hm. \tag{8.7}$$

m n'est pas une probabilité. Pour $a = -1, h = 0$, les points stationnaires satisfont $m(m^2 - 1) = 0$. Les minima sont ± 1, le maximum est zéro et la barrière vaut $1/4$. Il s'agit d'un système bistable illustratif, non d'une mesure de coopération du produit.

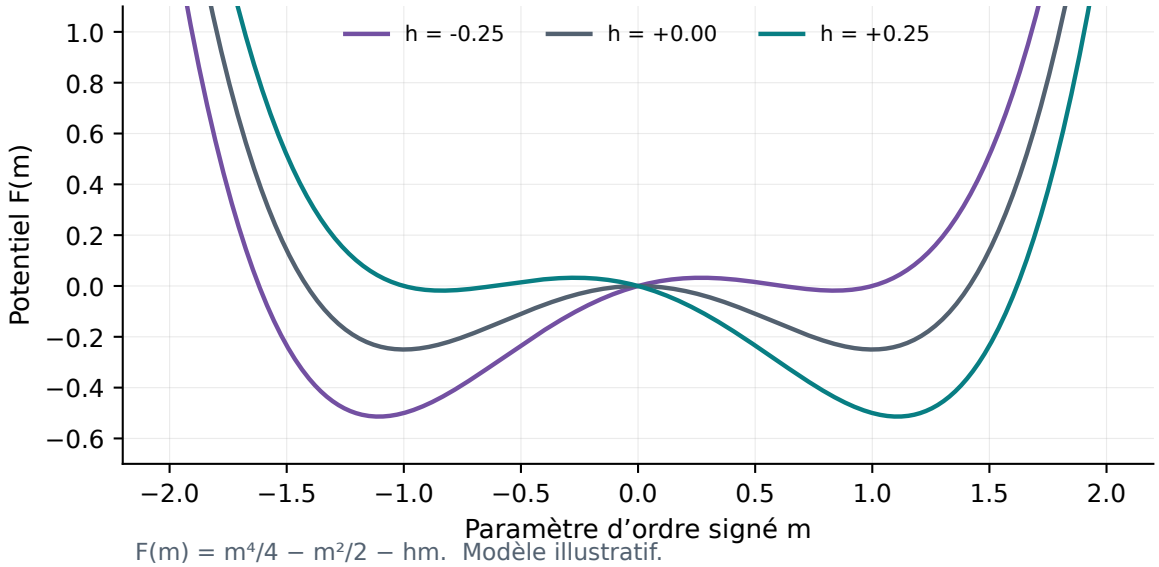


Figure 8.1 : Potentiels explicites $F(m) = m^4/4 - m^2/2 - hm$ pour trois biais. Un bassin peut représenter la coordination ou la capture selon l'interprétation de l'état. La stabilité ne prouve pas la légitimité.

8.7 Évasion et réparation contrôlée

Pour la diffusion à gradient

$$dm_t = -U'(m_t)\,dt + \sqrt{2D}\,dW_t, \tag{8.8}$$

l'approximation d'évasion à faible bruit entre puits non dégénérés possède une dépendance exponentielle

$$k \propto e^{-\Delta U/D}. \tag{8.9}$$

Le préfacteur, les frontières et la convention de bruit demeurent essentiels [20, 33]. Ce résultat ne chiffre pas le risque d'attaque d'une institution. L'analogie suggère des barrières élevées vers la capture, mais des chemins faisables vers la réparation, l'appel et la sortie.

8.8 Taille finie et diagrammes de phase

Pour un ensemble fini de coûts lisses et une température effective positive, la fonction de partition est une somme finie de termes positifs. Aucune singularité thermodynamique n'en découle. Un changement rapide, une bifurcation dynamique, un saut métastable et une discontinuité du minimum à bruit nul sont des phénomènes distincts.

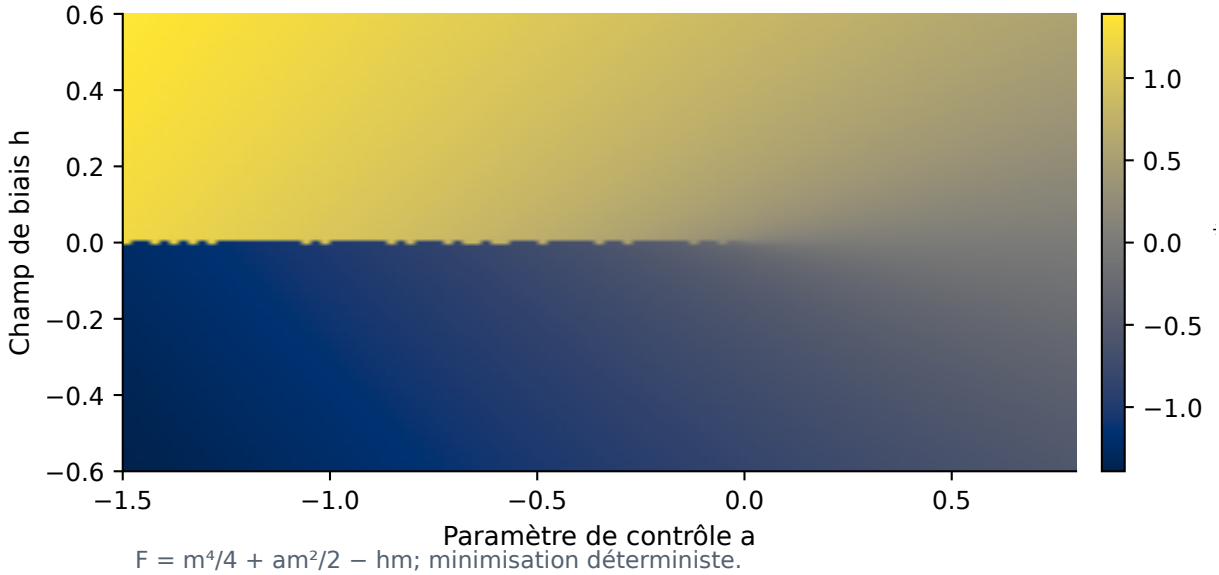


Figure 8.2 : Minimum déterministe de $m^4/4 + am^2/2 - hm$ sur une grille déclarée. Ce n'est pas un diagramme empirique des institutions. Pour $h = 0, a < 0$, deux minima sont égaux ; la grille en choisit un représentant.

Une hypothèse de transition institutionnelle exige un paramètre d'ordre observable, un graphe d'interactions et des tests contre d'autres explications. Dans un modèle échangeable à corrélation commune ρ, la taille équivalente en variance est $N_{\text{eff}} = N/[1 + (N-1)\rho]$. Le nombre d'adresses ne constitue pas le nombre de domaines indépendants de contrôle.

8.9 Gibbs et Helmholtz : conserver les unités

L'énergie libre physique de Gibbs est $G = H_{\text{enthalpie}} - TS$. La relation $\Delta G = \Delta H_{\text{enthalpie}} - T\Delta S$ suppose la même température aux deux états ; sinon il faut conserver $\Delta(TS)$. L'énergie de Helmholtz est $F = U - TS$ et la relation canonique $F = -k_B T \log Z$ exige un ensemble physique adéquat. Le coût régularisé de l'équation (8.3) n'est pas une énergie de Gibbs mesurée pour une institution.

Portée démontrée

L'identité variationnelle est prouvée dans son cadre. Les figures reproduisent des fonctions explicites. L'interprétation institutionnelle exige variables observables, unités, loi de transition estimée, incertitude et réfutation possible. Ni un faible coût modélisé ni un équilibre stable ne confère droits, vérité ou autorité.

9 Continuité hamiltonienne, information et responsabilité causale

9.1 La continuité comme flux contraint

La mécanique hamiltonienne décrit une évolution structurée dans l'espace des phases [26]. Dans le modèle institutionnel, q représente la mission, les rôles, la mémoire et la topologie ; p les engagements accumulés : contrats, obligations, capacités, dépendances et coûts de transition. Un Hamiltonien de base est

$$H(q, p, t) = T(p) + V(q, t), \tag{9.1}$$

avec

$$\dot{q} = \nabla_p H, \qquad \dot{p} = -\nabla_q H. \tag{9.2}$$

Le calcul canonique donne

$$\frac{dH}{dt} = \partial_t H. \tag{9.3}$$

La conservation exige donc un Hamiltonien autonome, ou une dérivée temporelle nulle le long de la solution. Les coordonnées conjuguées doivent être justifiées ; les obligations ne sont pas automatiquement des moments mécaniques. Il ne s'agit pas d'une conservation d'énergie physique. Le modèle souligne que la succession doit préserver certains invariants tout en exposant l'inertie cachée des engagements.

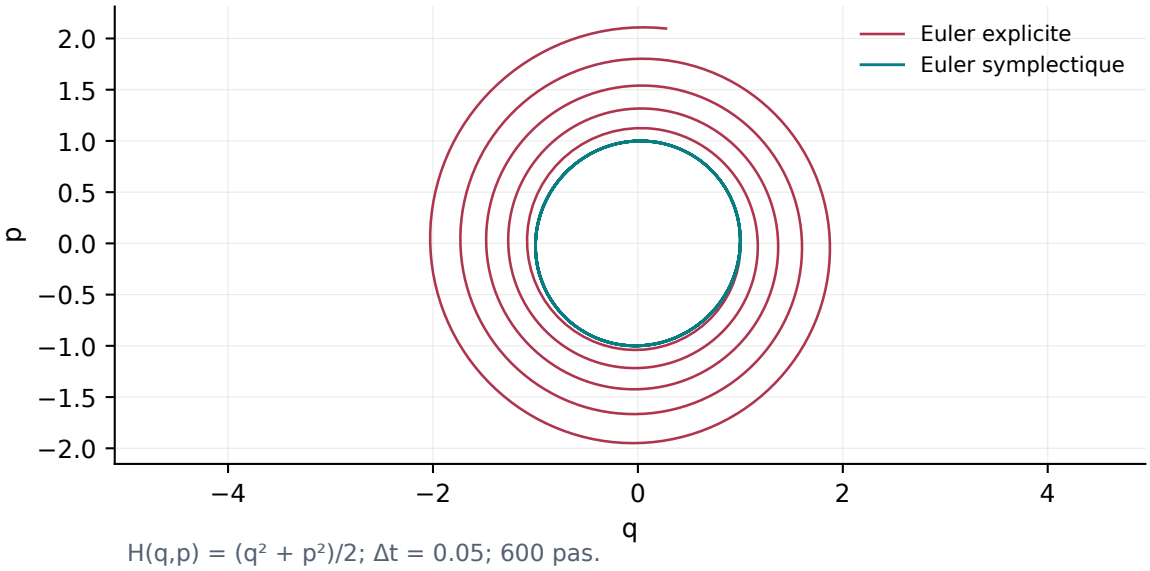


Figure 9.1 : Oscillateur explicite $H = (q^2 + p^2)/2$: Euler explicite et Euler symplectique, pas de 0.05 et 600 itérations. Une erreur d'énergie bornée dans cet exemple n'est pas une conservation exacte, ni une propriété démontrée de la Chronique.

9.2 Dynamique port-hamiltonienne

Les institutions réelles sont ouvertes, contrôlées et dissipatives :

$$\dot{z} = [J(z) - R(z)]\nabla H(z) + G(z)u + G_w(z)w, \tag{9.4}$$

avec $J = -J^\top$, $R = R^\top \succeq 0$, entrées u et perturbations w. Pour H autonome, posons $y = G^\top \nabla H$ et $y_w = G_w^\top \nabla H$. Alors

$$\dot{H} = -\nabla H^\top R \nabla H + y^\top u + y_w^\top w. \tag{9.5}$$

Cette identité de stockage n'établit pas la stabilité asymptotique. Il faut notamment des conditions d'équilibre, de positivité, de détectabilité et de rétroaction [43]. L'expiration, la révocation et la succession sont des sauts hybrides, pas automatiquement des dissipations continues. Le lien avec les budgets et la péremption exige une correspondance des variables et des unités. Les demandes des clients, résultats de preuve, changements de politique et règlements sont des ports qui exigent authentification et observabilité.

9.3 Variété constitutionnelle

Pour des contraintes $\phi_j(z) = 0$ et $g_k(z) \le 0$,

$$\mathcal{M} = \{z \,:\, \phi(z) = 0,\ g(z) \le 0\}. \tag{9.6}$$

Une mise à jour contrainte est

$$z_{t+1} = \Pi_{\mathcal{M}}(z_t + \Delta t\, f(z_t, u_t)). \tag{9.7}$$

L'ensemble n'est pas nécessairement une variété lisse. Une projection unique demande des hypothèses supplémentaires, par exemple un ensemble fermé, convexe et non vide. Une projection discrète ne prouve pas l'invariance continue. Un échec mène à l'abstention ou au rejet, non à un relâchement silencieux. Les invariants comprennent l'impossibilité pour le successeur d'effacer seul la Chronique, l'expiration de l'autorité, la séparation entre acceptation et preuve, la porte des droits et la requalification après modification matérielle.

9.4 Relecture de la Chronique

La conception événementielle évite d'écraser l'état. Une transition

$$X_{t+1} = \mathsf{T}(X_t, e_t) \tag{9.8}$$

doit être déterministe pour un événement canonique ou consigner explicitement la non-détermination. La relecture depuis un point de confiance devrait reconstruire le successeur, l'admission et l'autorité. Sinon, l'état public n'est pas auditable.

9.5 Qualité informationnelle de la preuve

Pour une capacité Y et une preuve E,

$$I(Y;E) = H(Y) - H(Y \mid E). \tag{9.9}$$

Un gros paquet de documents peut apporter peu d'information s'il répète des observations corrélées. Un test adverse étroit peut être plus discriminant. Le plan peut maximiser

$$T^* = \arg\max_T \frac{\mathbb{E}[I(Y;E_T)]}{C(T) + \lambda R(T)}. \tag{9.10}$$

La divergence

$$D_{\mathrm{KL}}(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \tag{9.11}$$

peut mesurer le déplacement entre l'environnement de preuve et celui du déploiement. Une dérive importante doit contracter l'autorité.

9.6 Responsabilité causale

Une corrélation entre un successeur et un bon résultat ne suffit pas. Un modèle causal distingue action, contexte, résultat et facteurs non observés [55]. L'effet

$$\mathbb{E}[Y \mid \mathrm{do}(A = a)] - \mathbb{E}[Y \mid \mathrm{do}(A = a')] \tag{9.12}$$

n'est identifiable que sous des hypothèses ou un dessin expérimental déclarés. La Chronique doit donc conserver l'action, la cible, la version, les approbations humaines, les changements externes et la méthode de mesure.

Un contrat de preuve causale spécifie l'estimande, l'intervention, le comparateur, les critères d'inclusion, les covariables, les hypothèses d'interférence, les données manquantes, le calendrier de mesure, la sensibilité et la décision d'autorité influencée.

9.7 Variété requise

La cybernétique rappelle qu'un régulateur doit posséder une variété suffisante pour absorber celle de l'environnement [4, 17, 65]. Cette variété ne vient pas seulement d'un plus grand modèle, mais de candidats multiples, d'humains, de vérificateurs externes, de preuves diverses, d'adaptateurs de politique et de modes de retour arrière.

L'architecture distribue donc la variété épistémique et centralise uniquement les invariants constitutionnels minimaux. La souveraineté devient administration responsable plutôt que commandement omniscient.

10 MCAP-001 et MCAP-002

PARTIE III

PROTOCOLES ET IMPLÉMENTATION INSTITUTIONNELLE

10.1 Deux protocoles, deux questions

SUCCESSOR Ω sépare l'adressabilité du pont institutionnel.

- **MCAP-001 — Protocole d'adressabilité des Civilisations-machines** répond à la question : depuis une racine persistante, comment un client indépendant découvre-t-il le Pacte, les rôles, le successeur, les preuves, l'autorité, la Chronique et les limites ?
- **MCAP-002 — Protocole du pont institutionnel** répond à la question : comment le travail et le règlement économiques deviennent-ils mémoire admissible, autorité bornée, retour arrière et succession qualifiée sans confusion des catégories ?

Un nom ENS peut fournir identité et routage lisibles [30] ; il ne démontre ni capacité, ni sûreté, ni droit, ni légitimité. MCAP-001 publie l'état découvrable. MCAP-002 gouverne les transitions conséquentes.

10.2 Découverte par la racine MCAP-001

Une racine conforme publie : une interface humaine canonique ; un Manifeste racine lisible par machine ; un Pacte versionné ; un registre de rôles et de lignée ; le statut actif, proposé, suspendu ou remplacé de chaque composant ; les surfaces de preuve, vérification, admission, autorité, retour arrière et Chronique ; ainsi que les limites et non-prétentions.

La racine de référence est asi.eth. Sa source publique décrit actuellement une publication Genesis et maintient une porte de signature du propriétaire avant que les enregistrements sur chaîne ne soient qualifiés d'actifs [46, 49]. La publication du code et la liaison sur Ethereum sont deux événements différents.

Test de la racine

ASI.eth fonctionne comme racine adressable lorsqu'un client indépendant peut partir du nom, résoudre le manifeste, reconstruire le registre des rôles et l'état du successeur, trouver preuves et autorité, et découvrir les limites sans instruction privée de MONTREAL.AI.

10.3 Espace de noms canonique

Table 10.1 : Sémantique des rôles sous ASI.eth.

Nom d'exemple	Rôle	Sens requis
asi.eth	Racine	Découverte et espace de noms ; aucune autorité implicite
successor.asi.eth	Successeur	Continuation qualifiée courante ou proposée
covenant.asi.eth	Pacte	Invariants constitutionnels stables
constitution.asi.eth	Constitution	Règles et schémas de politique versionnés
registry.asi.eth	Registre	Identités, missions, contrôleurs et lignée
chronicle.asi.eth	Chronique	Engagements d'événements institutionnels
proof.asi.eth	Preuve	Candidats figés, plans, preuves et résultats
verifier.asi.eth	Vérificateur	Découverte des vérificateurs indépendants
admission.asi.eth	Admission	Qualification, portée, échéance, suspension et révocation
policy.asi.eth	Politique	Contraintes externes et plafonds
rollback.asi.eth	Retour arrière	Procédures, déclencheurs et reçus de récupération
treasury.asi.eth	Trésorerie	Ressources bornées, séquestre et règlement

La séparation sémantique doit être appuyée par une séparation du contrôle. Un propriétaire bénéficiaire peut contrôler plusieurs portefeuilles ; le registre doit donc divulguer contrôleurs, signataires, incitations, conflits, accès à la preuve et voies de récupération.

10.4 Pont MCAP-002

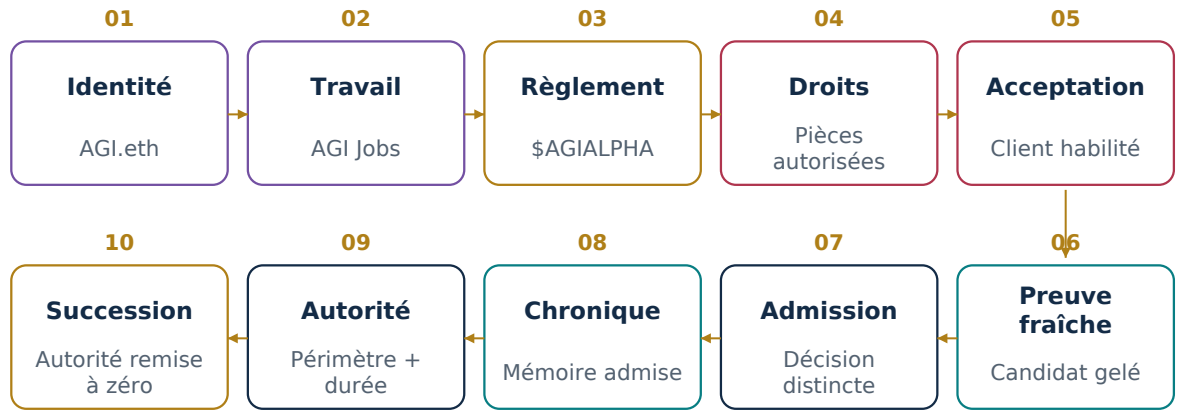
La chaîne complète est

$$\begin{aligned}\text{Identite AGI.eth} &\rightarrow \text{Mission AGI Jobs et sequestre}\\ &\rightarrow \text{Travail et elements probants}\\ &\rightarrow \text{Recu de reglement AGIALPHA}\\ &\rightarrow \text{Liberation des droits}\\ &\rightarrow \text{Acceptation du client}\\ &\rightarrow \text{Preuve independante}\\ &\rightarrow \text{Admission dans la Chronique ASI.eth}\\ &\rightarrow \text{Autorite bornee}\\ &\rightarrow \text{Surveillance, retour arriere, succession.}\end{aligned} \tag{10.1}$$

L'ordre évite les raccourcis. Le règlement peut précéder l'acceptation selon le code ; l'acceptation peut précéder la preuve générale ; la preuve peut établir une capacité sans autoriser une action.

10.5 Paquets requis

MCAP-002 requiert neuf types : reçu de règlement, manifeste de droits, reçu d'acceptation, reçu de preuve indépendante, décision d'admission, enveloppe d'autorité, événement de Chronique, reçu de révocation ou de retour arrière et dossier de succession. Chaque paquet référence les précédents sans les remplacer.



Actes distincts, pas un ordre universel : les droits peuvent précéder le travail.

Un règlement ne prouve ni les droits, ni l'acceptation, ni l'autorité.

Figure 10.1 : Chaîne de paquets MCAP-002. L'architecture n'est complète que lorsque chaque transition conséquente possède un enregistrement distinct et vérifiable.

10.6 Transitions interdites

Les promotions suivantes sont interdites : déclaration directement vers admission ; règlement directement vers preuve ; acceptation directement vers droits ; preuve directement vers autorité sans Principal ; nomination directement vers exploitation sans preuve fraîche ; autorité du prédécesseur directement vers autorité du successeur. Il

s'agit de règles de sûreté de type pouvant être appliquées par schémas, signatures, passerelles et relecture de la Chronique.

10.7 Adaptateurs d'interopérabilité

La doctrine stable demeure indépendante des interfaces changeantes. ENS fournit la découverte des noms et enregistrements [29, 30, 56, 57] ; MCP expose outils et ressources aux applications d'agents [41] ; A2A coordonne des tâches entre agents [1] ; ERC-8004 propose des registres émergents d'identité, réputation et validation [19] ; EIP-712 et ERC-1271 structurent les signatures [10, 22] ; Sign-In with Ethereum et les justificatifs vérifiables soutiennent l'authentification institutionnelle [15, 67].

Les standards à l'état de projet sont des cibles d'adaptation, non des dépendances constitutionnelles. L'adaptateur peut changer ; l'invariant de requalification demeure.

10.8 Langage normatif

Les termes DOIT, NE DOIT PAS, DEVRAIT et PEUT correspondent au langage normatif des standards Internet [12, 37]. La conformité porte sur des paquets et transitions observables, non sur des intentions.

11 Sémantique et canonicalisation des paquets

11.1 Identifiants canoniques

Chaque paquet possède un identifiant stable, par exemple

$$\mathrm{id} = \mathrm{urn} : \mathrm{mcap} : \mathrm{type} : \mathrm{reseau} : \mathrm{sujet} : \mathrm{version} : h. \tag{11.1}$$

Le condensat h repose sur une canonicalisation qui fixe l'ordre des champs, la normalisation Unicode, les nombres, le format temporel, les valeurs absentes et le domaine de signature. Un nom ENS peut résoudre vers l'identifiant sans devenir sa seule source de permanence.

11.2 Signatures typées

Une signature institutionnelle lie domaine, type de paquet, version de schéma, sujet, condensat, nonce, date d'émission et échéance. EIP-712 fournit un modèle de données typées ; ERC-1271 permet la validation par des comptes contrats [10, 22]. Le protocole doit encore établir la signification institutionnelle de la clé : contrôleur, rôle, révocation et conflits.

11.3 Validité et fermeture des références

La validité structurelle s'écrit

$$\mathsf{Valide}(p) = \mathsf{Schema}(p) \wedge \mathsf{Canonique}(p) \wedge \mathsf{Condensat}(p) \wedge \mathsf{Signature}(p) \wedge \mathsf{Frais}(p) \wedge \mathsf{References}(p). \tag{11.2}$$

Elle ne prouve pas la vérité substantielle. Une décision est fermée par référence seulement si toutes ses dépendances sont résolubles, versionnées et non révoquées. Une enveloppe dont la preuve ou le manifeste de droits a expiré doit être suspendue.

11.4 Statut et non-prétentions

Toute surface publique devrait publier état, classe de preuve, niveau d'autorité, dernière vérification, limites et prochaine porte. Ainsi, « actif » ne peut confondre publication, liaison ENS, preuve, adoption et exploitation.

12 AGI.eth, AGI Jobs et la preuve économique

12.1 La coordination économique comme engagement observable

AGI.eth constitue la racine des agents et de l'économie des machines. AGI Jobs fournit des contrats pour publier un travail, séquestrer des fonds, assigner un agent, consigner des points de contrôle, demander l'achèvement, valider, contester, régler, attribuer une réputation et émettre un reçu. AGIALPHA est l'actif de coordination des déploiements historiques Ethereum examinés ici [44, 45].

Le jeton n'est pas une preuve. Il rend les engagements mesurables par le séquestre et les cautions, mais un acteur riche peut rester incompétent, collusif ou non autorisé. Le rôle institutionnel est plus étroit : rendre explicites les enjeux, exposer des parties à des conséquences définies, rémunérer l'évaluation, conserver un règlement et émettre un reçu durable.

12.2 Éléments historiques sur Ethereum

Le registre v9 normalise trois emplois complétés par protocole :

Table 12.1 : Éléments normalisés de règlement AGI Jobs.

Variante	Gestionnaire	Emploi	Résultat consigné
Prime	`0xF8fc...993e`	0	Paiement principal de 100 000 AGIALPHA ; 80 000 observés à l'agent ; trois transferts aux validateurs ; NFT d'achèvement
Employer Burn	`0xBF66...F9eC2`	0	8 000 observés à l'agent ; trois validateurs ; NFT ; traitement du crochet ENS
Genesis	`0xB3AA...81d1`	8	400 000 observés à l'agent ; sept validateurs ; NFT d'achèvement

INSTANCE	JETONS À L'AGENT	CLASSE
Prime · Job 0	80,000 AGIALPHA	E1
Employer Burn · Job 0	8,000 AGIALPHA	E1
Genesis · Job 8	400,000 AGIALPHA	E1

Source : registre du produit v9. Aucun nouveau rejeu RPC dans cette révision.

FIGURE 12.1 : Trois règlements complétés par protocole. La promotion de E1 à E2 exige le rejeu indépendant des transactions.

Le dossier Prime Job 0 est le plus complet : création, désignation du gagnant, prise de la place, point de contrôle, demande d'achèvement et finalisation. Le règlement final est `0x5948...d69b` ; l'agent est associé à eva.agent.agi.eth ; le NFT est le jeton 0 du contrat Prime. Le crochet ENS a été ignoré parce qu'il n'était pas configuré.

Les dossiers Employer Burn et Genesis contiennent la finalisation et les observations économiques, mais pas encore tous les précurseurs. Ils restent E1. Un registre E2 devrait interroger deux fournisseurs Ethereum indépendants, décoder les journaux à partir du code et de l'ABI vérifiés, conserver le hash du bloc et distinguer la liaison ENS au moment de l'événement du nom inverse courant.

12.3 Ce que la boucle démontre déjà

Soutenu au niveau E1

- Des contrats AGI Jobs sur Ethereum ont atteint des états de règlement terminal.
- AGIALPHA a circulé sous forme de séquestre, cautions, paiements d'agents et transferts aux validateurs.
- Les exemples normalisés comprennent au moins trois reçus NFT d'achèvement.
- Plusieurs variantes de contrats ont démontré l'exécution protocolaire.
- Des identités ENS apparaissent dans l'historique opérationnel.

La maturité correcte est donc : déploiement principal complété ; premier règlement complété ; répétabilité multivariante démontrée ; rejeu indépendant, validation commerciale externe, assurance de sécurité et admission ASI.eth en attente.

Table 12.2 : Maturité de la boucle économique.

Jalon	Statut	Signification
Déploiement Ethereum	COMPLET	Contrats déployés et inspectables
Premier règlement AGIALPHA	COMPLET	Au moins un cycle économique terminal
Exécution multi-emplois et multivariante	DÉMONTRÉE	Trois variantes normalisées
Rejeu indépendant	EN ATTENTE	Paquet E2 à produire
Validation d'un client indépendant	EN ATTENTE	Aucun reçu public d'un client non apparenté dans cet article
Assurance de sécurité indépendante	EN ATTENTE	La source vérifiée n'est pas un audit
Admission institutionnelle ASI.eth	EN ATTENTE	Les règlements ne sont pas encore mémoire MCAP admise

12.4 Compilateur de reçu de règlement

Le compilateur doit résoudre la chaîne, le gestionnaire, la variante et l'ABI ; récupérer les reçus ; décoder les événements et transferts ; réconcilier séquestre, cautions, paiements, récompenses et soldes ; consigner l'identité au moment de l'événement ; résoudre les URI ; identifier le NFT ; consigner le crochet ENS ; et calculer un condensat canonique.

L'invariant comptable générique est

$$E_0 + B_A + \sum_v B_v + B_D = P_A + P_V + R_E + R_B + R_P + \Delta L, \tag{12.1}$$

avec séquestre, cautions, paiements, remboursements, montant retenu et solde verrouillé expliqué. Un reçu qui ne se réconcilie pas doit être qualifié d'inconclusif.

12.5 Identité au moment de l'événement

Les noms ENS sont mutables. Le reçu conserve donc adresse, engagement de nom au temps t, nom résolu lors de la vérification, résolveur, bloc et méthode. L'adresse et le journal du bloc constituent la preuve économique primaire ; le nom ajoute la sémantique seulement lorsque sa liaison temporelle est établie.

12.6 Sémantique du NFT d'achèvement

Un NFT peut servir de reçu s'il lie gestionnaire, emploi, transaction, employeur, agent, preuve d'achèvement, flux économiques, version du protocole, limites et classe probante. Son transfert ne transfère ni droits de propriété intellectuelle, ni autorité institutionnelle, sauf instrument distinct.

12.7 Du règlement à la mémoire

Un reçu protocolaire devient admissible seulement après revue des droits et de la pertinence. Les événements historiques peuvent entrer dans la Chronique comme

$$\text{REGLEMENT\_IMPORTE}[\text{pre-MCAP}, E1, \text{aucune autorite}]. \tag{12.2}$$

La promotion ultérieure vers E2 conserve l'événement initial et ajoute le rejeu indépendant.

Conclusion économique corrigée

AGI.eth, AGI Jobs et AGIALPHA ont déjà exécuté une boucle économique réelle au niveau protocolaire. Le prochain seuil consiste à démontrer qu'un client non apparenté peut s'y fier de manière répétée pour un travail libéré de droits, accepté, vérifié indépendamment et intégré à une succession gouvernée.

13 Droits, acceptation du client et vérification indépendante

13.1 Pourquoi le règlement n'est pas l'acceptation

Un contrat intelligent peut exécuter une transition sans savoir si le livrable hors chaîne est licite, utile, complet ou accepté par un client autorisé. Le pont institutionnel conserve donc quatre questions distinctes : le protocole a-t-il réglé ? Les entrées, outils, données et sorties ont-ils été utilisés sous des droits valides ? Un client autorisé a-t-il accepté le résultat selon des critères préengagés ? Un vérificateur indépendant a-t-il établi la prétention plus large sur un candidat figé ?

Les réponses peuvent diverger. Le code peut régler un emploi alors que le client rejette ensuite le livrable. Le client peut accepter un résultat utile tandis qu'une licence de données reste incertaine. Un résultat accepté et libéré de droits peut échouer à un test de généralisation. L'autorité doit suivre la porte pertinente la plus faible.

13.2 Manifeste de libération des droits

Un manifeste s'écrit

$$\begin{aligned}\mathsf{R} = \langle &\text{mission}, \text{entrees}, \text{donnees}, \text{modeles}, \text{outils}, \text{sorties},\\ &\text{juridictions}, \text{restrictions}, \text{conservation}, \text{signataires}, \text{echeance}\rangle.\end{aligned} \tag{13.1}$$

Pour chaque actif a,

$$D(a) \in \{\text{propre}, \text{licencie}, \text{autorise}, \text{domaine public}, \text{restreint}, \text{non resolu}\}. \tag{13.2}$$

La mémoire exige un état compatible avec l'usage proposé. « Accessible publiquement » n'est pas une base de droits. Le manifeste consigne source, licence ou contrat, finalité permise, portée territoriale et temporelle, œuvres dérivées, attribution, confidentialité, données personnelles et révocation.

Porte dure des droits

Aucun résultat ne peut devenir mémoire admise, preuve publique ou héritage d'un successeur tant qu'une dépendance matérielle de droits demeure non résolue. L'élément peut rester en quarantaine avec son histoire.

Le manifeste structure la décision de l'avocat, du client et de l'opérateur ; il ne remplace pas leur jugement.

13.3 Reçu d'acceptation du client

Un reçu d'acceptation contient

$$\begin{aligned}\mathsf{RAC} = \langle &\text{client}, \text{pouvoirDeSigner}, \text{mission}, \text{criteres}, \text{denominateur},\\ &\text{resultat}, \text{exceptions}, \text{effetEconomique}, \text{droits}, \text{signature}, \text{temps}\rangle.\end{aligned} \tag{13.3}$$

Les critères préengagés empêchent de redéfinir le succès après observation. Le dénominateur complet inclut les cas rejetés, abstentions et exceptions. Le reçu distingue livraison, conformité technique, résultat d'affaires, autorisation de divulgation, admission en mémoire et permission d'entraînement futur. Une seule signature n'autorise pas nécessairement tous ces usages.

13.4 Indépendance du vérificateur

L'indépendance est relative à une prétention et à un modèle de menace. Définissons

$$\iota(V, C) = \big(i_{\text{controle}}, i_{\text{finance}}, i_{\text{preuve}}, i_{\text{methode}}, i_{\text{incitation}}, i_{\text{appel}}\big). \tag{13.4}$$

Un score agrégé peut aider au triage, mais certains conflits sont des échecs durs. Si le demandeur peut modifier la preuve décisive après le gel, l'indépendance probante est nulle même si d'autres composantes sont élevées.

La divulgation minimale couvre contrôle légal et bénéficiaire, rémunération, relation antérieure, accès aux données brutes, auteur du plan, pouvoir de modifier les métriques, garde des clés, révocation, appel et publication.

13.5 Marché de la vérification

Un marché accroît la capacité, mais crée une sélection adverse : les vérificateurs faibles peuvent sous-évaluer la diligence et les demandeurs choisir le plus permissif. Les protections comprennent affectation aléatoire ou réglementée, filtrage des conflits, rémunération indépendante du résultat, cautions liées aux violations de processus, défis reproductibles, réputation de calibration et droit d'appel.

Lorsque les résultats deviennent observables, une règle de score propre telle que

$$S(p, Y) = Y \log p + (1 - Y) \log(1 - p) \tag{13.5}$$

peut récompenser la calibration. Elle ne remplace pas la preuve lorsqu'aucune vérité finale observable n'existe.

13.6 Agrégation sans blanchir l'incertitude

Si les vérificateurs fournissent des rapports de vraisemblance L_v, une agrégation prudente est

$$\log L_{\text{agg}} = \sum_v w_v \log L_v, \qquad \sum_v w_v \le 1, \tag{13.6}$$

avec poids réduits en présence de sources ou de contrôleurs communs. Le reçu conserve le désaccord au lieu de produire un nombre unique qui cache l'incertitude.

13.7 Preuve publique

Une carte de preuve publique, générée seulement à partir d'un sous-ensemble divulguable et libéré de droits, peut présenter mission, versions, protocole d'évaluation, dénominateur, résultats, incertitude, acceptation, vérificateur, conflits, autorité, limites, échéance et engagements de Chronique. Elle ne publie pas nécessairement les données protégées.

13.8 Passage vers E4

La promotion exige

$$E4 \Leftarrow E3 \wedge \mathsf{DroitsLiberes} \wedge \mathsf{ClientAccepte} \wedge \mathsf{ConsequenceBornee} \wedge \mathsf{ResultatObserve} \wedge \mathsf{Chronique}. \tag{13.7}$$

E4 ne signifie ni sûreté universelle ni confiance permanente. Il désigne une prétention précise ayant produit une conséquence mesurable et gouvernée. Tout changement matériel ou toute expiration exige requalification.

Doctrine institutionnelle

Le règlement dit ce que le code a fait. Les droits disent ce que les parties pouvaient faire. L'acceptation dit ce que le client a accepté. La preuve indépendante dit ce que soutiennent les éléments probants. L'admission dit ce que l'institution décide. L'autorité dit ce que le successeur peut faire.

14 Sécurité, déploiement et contrôle opérationnel

14.1 Modèle de menace

L'objectif de sécurité dépasse l'appel non autorisé à une API. Il faut empêcher un acteur de fabriquer une légitimité institutionnelle ou de conserver une autorité après disparition de ses conditions. Les menaces comprennent auto-certification, capture du vérificateur, substitution de preuve, compromission de clés, résolution ENS périmée, rejeu de paquets, fausse déclaration de droits, usurpation du client, dérive des privilèges, action irréversible, réécriture de Chronique, drainage de trésorerie et dérive du modèle ou des données.

14.2 Frontières de confiance

Une implémentation complète distingue écriture de la mission, génération du candidat, stockage figé, collecte de preuve, vérification indépendante, acceptation, admission, passerelle d'action, garde de Chronique, trésorerie et retour arrière. Pour A0, certaines fonctions peuvent être co-localisées avec divulgation ; pour A2 et plus, un administrateur unique contrôlant tout invalide la prétention d'indépendance.

14.3 Autorisation zéro confiance

Les principes zéro confiance exigent authentification et autorisation explicites pour chaque ressource et action [14, 58]. Une requête contient sujet, enveloppe, condensat d'implémentation, action, cible, contexte, nonce, temps et signature. La passerelle vérifie le sujet, l'état des justificatifs, la portée, les limites, l'échéance, les approbations, les dépendances de preuve et de droits, la viabilité, l'idempotence, l'observabilité et le retour arrière.

L'échec d'une porte dure réduit l'autorité ou refuse l'action. La passerelle ne demande pas au modèle s'il croit être autorisé.

14.4 Profils A0, A1 et A2

A0 observe, résout, simule et recommande sans identifiant d'action externe. A1 émet une recommandation attribuable à un humain responsable, avec preuves et incertitude. A2 agit dans un bac à sable sans voie vers les données ou identifiants de production. La référence v9 plafonne volontairement l'autorité à A2 et désactive les effets externes conséquents [47, 48].

A3 exige identité de production, moindre privilège, cibles et montants limités, simulation, idempotence, surveillance indépendante, retour arrière testé, intervention d'urgence, expiration de preuve et revue juridique. A4 est exceptionnel : action conséquente étroite, explicitement autorisée et de courte durée.

14.5 Intégrité de la chaîne d'approvisionnement

Une version vérifiable comprend manifeste, inventaire SHA-256, nomenclature logicielle, instructions, tests, provenance et script de vérification. SLSA et in-toto offrent des modèles utiles [53, 63]. Le condensat du sujet couvre source, dépendances, modèles, invites, configuration, schémas, politiques et construction. Toute modification matérielle produit une nouvelle version et déclenche la requalification.

14.6 Gouvernance des clés et des résolveurs

La propriété des racines, la gestion des enregistrements, la signature opérationnelle, la vérification et la trésorerie doivent utiliser des contrôles distincts proportionnels au risque. Les portefeuilles matériels, les périodes de renouvellement longues, les verrous de registraire, la répétition de récupération et la simulation des transactions réduisent le risque existentiel.

Les fuses ENS irréversibles ou les résolveurs verrouillés ne devraient être utilisés qu'après revue indépendante des hypothèses de contrôle, expiration, migration, récupération et urgence.

14.7 Surveillance et dérive

Pour une distribution de preuve P_0 et de déploiement Q_t, une statistique de dérive est

$$D_t = D_{\mathrm{KL}}(Q_t \| P_0). \tag{14.1}$$

L'autorité peut demeurer intacte sous δ_1, être rétrécie entre δ_1 et δ_2, puis devenir vide au-delà. Les seuils doivent être calibrés. La surveillance couvre erreurs, abstentions, exceptions, outils, dépenses, latence, remplacements humains et incidents.

14.8 Résilience opérationnelle

Les autorités financières et les opérateurs d'infrastructures critiques mettent l'accent sur le risque technologique, les tiers, la cybersécurité et la résilience des systèmes agentiques [52]. SUCCESSOR Ω rend ces exigences adressables : quel successeur a agi, sous quelle mission, preuve, autorité et durée ; qui a accepté et vérifié ; quelle cible a changé ; comment l'effet peut-il être concilié ou renversé ; quel événement est entré dans la Chronique ?

Conclusion de sécurité

La vérification de source, les condensats et les tests réussis sont nécessaires, mais ne constituent pas une assurance de sécurité indépendante. La production exige modèle de menace, revue externe, correction,

nouvel essai, identité et secrets, staging, test d'intrusion, récupération réelle et déploiement borné par les conséquences.

15 Cas I : intégrité des dépenses fournisseurs

PARTIE IV

CAS INSTITUTIONNELS DÉTAILLÉS

15.1 Pourquoi cette mission est appropriée

L'intégrité des dépenses fournisseurs constitue une première mission institutionnelle forte : valeur économique mesurable, périmètre borné et compréhension immédiate par la direction. Un agent peut détecter factures en double, prix hors contrat, conditions anormales, fractionnement d'achats, concentration des fournisseurs et incohérences fiscales ou monétaires sans transférer lui-même des fonds. L'autorité initiale demeure A1 : recommander une revue, un gel ou une correction à un analyste responsable.

La mission ne devrait pas être « minimiser les dépenses », objectif susceptible de détruire résilience, qualité ou équité. Une constitution plus robuste est :

> Détecter, sur la base de preuves, les occasions de prévenir ou récupérer des dépenses fournisseurs non autorisées, tout en préservant les droits contractuels, la continuité opérationnelle, l'auditabilité, la proportionnalité de la revue et l'autorité humaine sur les décisions de paiement conséquentes.

15.2 Constitution et fonction de valeur

Pour un paiement candidat i, avec montant v_i, probabilité estimée de perte évitable p_i, coût de revue c_i, coût de faux positif f_i et risque d'irréversibilité r_i,

$$J_i = p_i v_i - (1 - p_i) f_i - c_i - \lambda r_i. \tag{15.1}$$

Le système recommande une revue lorsque $J_i > \tau$ et que les portes de politique passent. Il ne bloque pas le paiement à A1. À A2, il peut reproduire le gel dans un environnement miroir. Une autorité A3 exigerait preuve propre au client, revue juridique, listes de cibles, plafonds de montant, expiration, surveillance et retour arrière.

Table 15.1 : Éléments de mission.

Élément	Exemple	Contrôle institutionnel
Bénéficiaires	Équipe financière, fournisseurs, actionnaires et obligations publiques	Analyse explicite des parties prenantes
Courant	Règles et contrôles ERP existants	Période et coût figés
Bêta	Processus loué fondé sur un modèle de pointe	Même accès aux données et outils
Candidat	Processus SUCCESSOR Ω appartenant au client	Condensat figé avant preuve
Valeur	Dépenses non autorisées évitées ou récupérées	Méthode monétaire validée par le client
Sûreté	Aucun refus de paiement sans décision humaine autorisée	Enveloppe A1 et contrôle d'interface
Droits	Factures, contrats et données fournisseurs autorisés	Manifeste de droits et confidentialité
Preuve	Cas frais, doubles adverses et risques de queue	Vérificateur indépendant et dénominateur complet
Retour arrière	Retrait de recommandation, version antérieure, quarantaine de mémoire	Reçu et événement de Chronique

15.3 Compétition des candidats

La Fonderie Alpha génère plusieurs programmes du monde : détection déterministe de doubles, graphes de fournisseurs et comptes, recherche contrat-facture, priorisation causale, critique en ensemble et bouclier de politique. Les candidats sont comparés au Courant et au meilleur Bêta sous accès égal. L'avantage de mission doit rester positif après coûts, risques, fragilité et transition.

15.4 Plan de preuve

Le plan fige candidat, schéma, politique et métriques ; définit le dénominateur complet ; sépare entraînement, calibration et cas frais ; inclut quasi-doubles, paiements récurrents légitimes, notes de crédit, change, taxes et changements de fournisseurs ; mesure précision, rappel, précision monétaire, temps d'analyste, abstention, violation de politique et perte de queue ; compare Courant, Bêta et candidat ; confie le rejeu à un vérificateur indépendant ; documente droits, acceptation, exclusions et incertitude.

La précision monétaire est

$$\mathrm{PM} = \frac{\sum_{i \in \text{vrais positifs}} v_i}{\sum_{i \in \text{signales}} v_i}. \tag{15.2}$$

Une forte précision événementielle peut gaspiller du temps sur de faibles montants ; un seul faux positif important peut détériorer une relation fournisseur.

15.5 Enveloppe A1

L'enveloppe lie le condensat du sujet, la mission, l'action « produire une recommandation de revue », une file de travail déclarée, l'interdiction de modifier l'ERP ou de contacter le fournisseur, des plafonds journaliers, une durée

de trente jours, la surveillance, la révocation, le retour arrière et les références de preuve. Toute modification matérielle provoque expiration ou requalification.

15.6 Pont économique

Une partie de la mission peut être publiée dans AGI Jobs. Des agents concourent, des validateurs examinent, AGIALPHA règle le résultat protocolaire. Le reçu du client reste séparé. Après droits, acceptation et preuve, ASI.eth peut enregistrer l'issue dans la Chronique et la rendre admissible à la mémoire du successeur.

Le solde de jetons peut déterminer l'admissibilité ; il ne prouve ni compétence ni autorité.

15.7 Critère E4

Le cas atteint E4 seulement si un client non apparenté accepte une issue libérée de droits, qu'un vérificateur réellement indépendant signe une preuve reproductible, qu'une enveloppe A1 ou A2 est appliquée, qu'une conséquence mesurable est observée et que la Chronique conserve la transition.

16 Cas II : qualification d'un successeur

16.1 Frontière générationnelle

Supposons qu'un prédécesseur Θ_0 exploite une mission à A1 et possède une mémoire admise. Le candidat Θ_1 utilise un autre fournisseur de modèle et un nouveau composant de raisonnement par graphe. Il reçoit

$$\mathcal{I}_0 = \langle \mathcal{M}_0, \mathcal{K}_0^{\mathrm{adm}}, \mathcal{O}_0, \mathrm{Refs}(\mathcal{C}_0) \rangle, \qquad \mathcal{A}_1^{\text{avant preuve}} = \varnothing. \tag{16.1}$$

16.2 Preuve fraîche

Le plan vérifie l'intégrité et les droits de la mémoire héritée ; retire les éléments périmés ; fige le condensat du successeur ; rejoue les cas du prédécesseur ; ajoute des cas de dérive ; teste les nouveaux risques de prompt, outils et graphes ; compare prédécesseur, Bêta et successeur ; exerce le retour arrière ; publie le résultat et les limites.

Le successeur n'a pas à prouver une supériorité universelle, mais un avantage propre à la mission et l'absence de régression dans la portée d'autorité demandée.

16.3 Requalification de l'autorité

Si Θ_1 améliore le rappel mais calibre moins bien les factures en devises, le Principal peut admettre une enveloppe plus étroite :

$$\mathcal{A}_1 = \mathcal{A}_0 \setminus \{\text{recommandations en devises au-dessus de } v^*\}. \tag{16.2}$$

Un système meilleur en moyenne peut mériter une autorité plus faible sur une région précise.

16.4 Séquence de Chronique

La Chronique consigne nomination, proposition d'héritage, gel, début de preuve, résultat avec limites, admission, émission d'autorité, remplacement du prédécesseur et expiration ou requalification ultérieure. Le prédécesseur demeure découvrable. La restauration de son code ne réactive pas automatiquement son autorité.

16.5 Chemin d'échec

En cas d'échec, l'institution peut conserver le prédécesseur sous une enveloppe encore valide, retourner à A0, réparer et renommer un candidat, forker l'implémentation ou mettre en quarantaine la mémoire impliquée. L'échec devient connaissance institutionnelle lorsqu'il demeure visible avec les correctifs.

Critère de qualification

Un successeur est qualifié lorsqu'un observateur indépendant peut reconstruire ce qu'il a hérité, ce qui a changé, comment il a été évalué, quelles limites demeurent, qui l'a admis, quelle autorité il a reçue et comment cette autorité expire ou est révoquée.

17 Conformité et reconnaissance institutionnelle

PARTIE V

CONFORMITÉ, LIMITES ET RECHERCHE

17.1 Niveaux de conformité

La conformité est une affirmation probante. MCAP-002 définit cinq niveaux :

Table 17.1 : Échelle de conformité MCAP-002.

Niveau	Éléments requis	Affirmation permise
Préparé	Schémas, modèles, validation locale et référence bornée	Architecture préparée ; aucune prétention indépendante
Lié à la preuve	Une chaîne complète reproductible par un tiers	Une chaîne précise peut être reconstruite
Opérationnel-A2	Client non apparenté, vérificateur indépendant, bac à sable et retour arrière	Exploitation bornée en bac à sable
Successeur qualifié	Héritage de mémoire, preuve fraîche et nouvelle autorité	Invariant de succession démontré
Répétable	Cycle complet dans au moins deux contextes non apparentés	Pont institutionnel répétable dans la portée déclarée

Le paquet v9 est conçu pour établir localement le niveau Préparé. Les règlements historiques soutiennent l'exécution protocolaire, mais ne promeuvent pas la chaîne entière. La reconnaissance commence lorsque des tiers peuvent résoudre, reproduire, contester, implémenter, citer et utiliser l'architecture.

17.2 Test d'aptitude à la finalité

L'objectif élargi exige qu'un client indépendant résolve ASI.eth ; reconstruise un règlement AGI Jobs ; trouve une disposition signée des droits ; constate l'acceptation d'un client autorisé non apparenté ; reproduise une preuve par un vérificateur indépendant ; constate une admission par un Principal séparé ; observe l'exploitation dans l'enveloppe et un retour arrière ; retrouve la mémoire admise dans une Chronique non effaçable par le successeur ; constate qu'un nouveau successeur gagne une nouvelle autorité ; et voit le cycle répété dans un second contexte.

17.3 Suite de tests

Les essais devraient couvrir schémas, canonicalisation, signatures, rejeu, fermeture des références, révocation, confinement d'autorité, mutation après gel, dénominateur incomplet, droits non résolus, conflits de contrôle, relecture et modification de Chronique, retour arrière, réinitialisation de l'autorité et reconstruction de bout en bout. Les résultats indiquent version, environnement, échecs, essais omis et limites.

17.4 Reconnaissance

La reconnaissance n'est pas auto-attribuée. Elle se manifeste par résolution indépendante, critique ou citation, implémentation externe, participation d'un vérificateur, client non apparenté, dépendance récurrente, références dans les standards ou l'approvisionnement et succession observée. Un rapport public d'échec corrigé peut avoir davantage de valeur qu'une publicité non inspectable.

18 Limites et falsifiabilité

18.1 L'architecture n'est pas l'exploitation

Un article, un schéma, un runtime local ou un espace de noms publié peut établir une préparation. Il ne peut créer un client non apparenté, un vérificateur indépendant, des droits juridiques, une assurance de sécurité ou une autorité à conséquences. Les portes ouvertes comprennent la liaison signée d'ASI.eth, le rejeu E2 des règlements, l'acceptation externe, la libération des droits, la garde indépendante de la preuve, l'audit de sécurité, le fonctionnement A1/A2 réel, la succession qualifiée et la répétition externe.

18.2 Limites statistiques et causales

La validité dépend de la distribution, de la mesure et des hypothèses causales. Un candidat peut réussir un jeu d'essai et échouer sous dérive. Le client peut être biaisé. L'indépendance peut être partielle. L'effet causal peut être non identifiable. Une prétention de « Specialist ASI » ne peut être que spécifique à une mission, un comparateur, un coût, un risque et une preuve ; elle n'implique aucune intelligence universelle.

18.3 Limites économiques

Les jetons peuvent coordonner les ressources ou concentrer le pouvoir, subventionner la manipulation et créer une ambiguïté réglementaire. Le cautionnement n'est utile que si les violations sont observables et contestables. Le prix du jeton n'est pas la qualité institutionnelle ; la solvabilité du contrat ne prouve pas l'équité hors chaîne.

18.4 Limites de gouvernance

La séparation des noms, portefeuilles et services ne prouve pas la séparation du contrôle bénéficiaire. La divulgation améliore la responsabilité, mais peut créer des risques de confidentialité. Un parent contrôlé par le fondateur préserve la continuité et peut aussi concentrer le risque. L'architecture atténue ces problèmes sans les abolir.

18.5 Limites des modèles physiques

Les modèles d'énergie libre, de transition de phase, hamiltoniens et de viabilité n'ont de valeur que si variables et paramètres sont opérationnalisés. Les systèmes sociaux peuvent modifier leurs règles, interpréter les preuves et contester les objectifs. Les modèles doivent produire des hypothèses vérifiables, non des affirmations métaphysiques.

18.6 Conditions de réfutation

La thèse serait affaiblie si l'autorité réinitialisée ne protège rien parce que les anciens identifiants restent actifs ; si la Chronique ne détecte pas la réécriture ; si la preuve indépendante coûte davantage que sa valeur ; si la séparation diminue la fiabilité sans réduire la capture ; si le client ne peut exercer la portabilité ou la sortie ; si la succession perd la mémoire utile ; si les incitations favorisent la manipulation ; ou si MCAP ne peut être implémenté indépendamment de MONTREAL.AI.

19 Programme de recherche

Les priorités comprennent la vérification formelle de l'invariant d'autorité, la fermeture des références, le confinement des enveloppes, la relecture de Chronique et les transitions interdites ; la conception de marchés de vérification résistants aux coalitions ; l'identification empirique de paramètres d'ordre et de barrières institutionnelles ; la preuve causale de mission ; la preuve préservant la confidentialité ; les institutions multijuridictionnelles ; et le pluralisme des Pactes et forks paisibles.

Les expériences décisives suivantes sont le rejeu indépendant des trois règlements, l'activation résolue d'ASI.eth, une mission A1 avec client non apparenté, une preuve externe avec limites, un retour arrière mesuré, une requalification de successeur, la répétition dans un second contexte et une implémentation MCAP indépendante. Chaque expérience devrait publier les transitions réussies et échouées.

20 Résultats analytiques et tests reproductibles

20.1 Portée

Ce chapitre expose les modèles et contre-exemples qui rendent les hypothèses vérifiables. Les résultats sont des conséquences algébriques ou des calculs déterministes sur des systèmes synthétiques. Ils ne constituent ni audit indépendant de sécurité, ni acceptation client, ni mesure d'intelligence, ni certification de production. Le code des figures, les données CSV et les tests logiques accompagnent le texte.

20.2 Admission comme contrôle contraint

Soit $U_k(x)$ l'ensemble des actions contextualisées permises par la contrainte k :

$$U_{\mathrm{adm}}(x) = \bigcap_{k \in \mathcal{G}} U_k(x), \qquad \mathcal{G} = \{\text{preuve, droits, mission, politique, risque, reprise, physique}\}. \tag{20.1}$$

Une nouvelle preuve peut élargir le premier ensemble sans supprimer les autres contraintes. Si l'intersection est vide, aucune solution autorisée n'existe. Pour un contrôle continu et $W \succ 0$, un filtre à intervention minimale est

$$u^*(x) = \arg \min_{u \in U_{\mathrm{adm}}(x)} \tfrac{1}{2} \|u - u_{\mathrm{nom}}(x)\|_W^2. \tag{20.2}$$

Un ensemble faisable non vide, fermé et convexe donne un minimum unique. La convexité stricte ne prouve pas la faisabilité. Les outils discrets, permissions non convexes et effets externes latents exigent un modèle adapté. L'abstention d'autorisation n'est pas nécessairement un contrôle physiquement stabilisant.

20.3 Budget de preuve pour les défaillances rares

Supposons un candidat gelé et un plan fixe de n essais de Bernoulli indépendants et identiquement distribués. Après zéro défaillance, la borne supérieure unilatérale exacte au niveau $1 - \alpha$ est

$$p_U = 1 - \alpha^{1/n}. \tag{20.3}$$

Elle résout $(1 - p_U)^n = \alpha$, cas particulier des limites binomiales exactes [16]. Pour $p_U \le \varepsilon$,

$$n \ge \left\lceil \frac{\log \alpha}{\log(1 - \varepsilon)} \right\rceil. \tag{20.4}$$

À 95% et $\varepsilon = 10^{-3}$, il faut au moins 2 995 essais sans défaillance. Il s'agit d'une couverture fréquentiste, non d'une probabilité a posteriori de sûreté. La conclusion reste propre à la distribution et aux dépendances supposées.

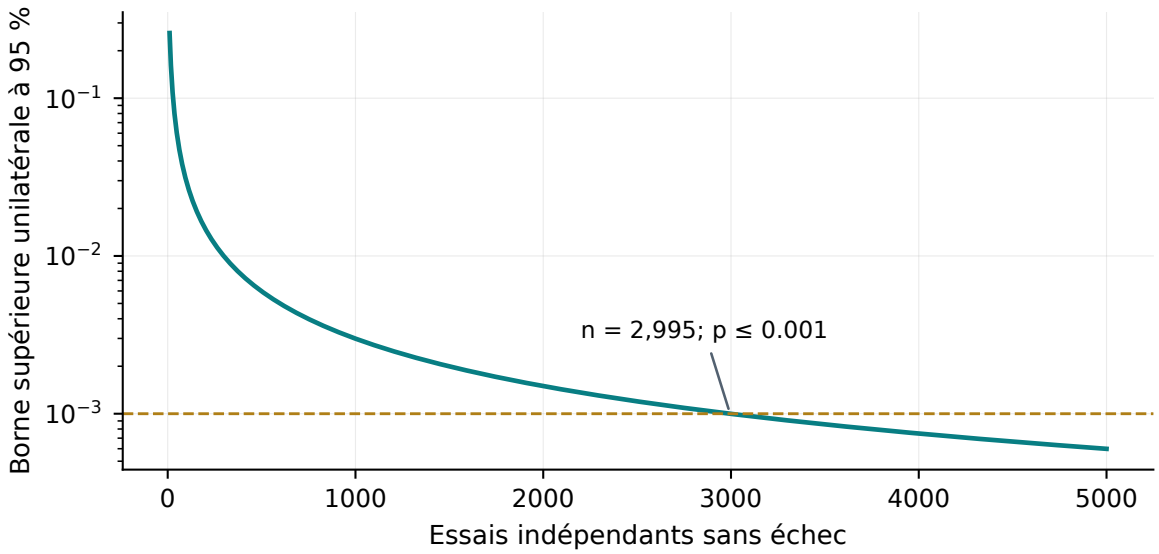


Figure 20.1 : Borne binomiale exacte $1 - 0.05^{1/n}$ sous échantillonnage indépendant à plan fixe. Une petite démonstration ne garantit pas un faible risque dans tous les déploiements.

La sélection parmi plusieurs candidats, l'ajustement sur le test final ou un arrêt opportuniste peuvent invalider cette couverture. Le plan doit fixer exclusions, abstentions et arrêt. Pour m prétentions distinctes, une allocation familiale simple utilise $\alpha_i = \alpha/m$. Les méthodes séquentielles exigent leur propre construction valide. Les cas manquants ne sont pas des succès.

20.4 Dissuasion, vérification imparfaite et fausses accusations

Pour un vérificateur neutre au risque, notons r la rémunération, $c \geq 0$ le coût de diligence, $g \geq 0$ le gain privé de collusion ou de négligence, $S \geq 0$ une sanction exécutoire, d la probabilité de détection et f celle d'une sanction erronée :

$$U_H = r - c - fS, \qquad U_D = r + g - dS. \tag{20.5}$$

La diligence est compatible avec les incitations exactement lorsque

$$(d - f)S \geq g + c. \tag{20.6}$$

La participation exige aussi $r \geq c + fS + u_0$ pour une option extérieure u_0. Si $d \leq f$ et $g + c > 0$, aucune sanction non négative ne suffit. Augmenter les cautions sans améliorer la discrimination peut décourager les participants honnêtes sans empêcher l'abus.

La recouvrabilité, l'adjudication, la responsabilité limitée, le contrôle commun et l'appel modifient le jeu. Un solde de jetons ne prouve aucune de ces propriétés. Dans un modèle de déclenchement répété particulier, la déviation d'une période est dissuadée si

$$p_d \left[S + \frac{\delta(R - P)}{1 - \delta} \right] \geq T - R, \tag{20.7}$$

avec détection p_d, tentation $T - R$, sanction crédible S, punition $P < R$ et $0 \leq \delta < 1$. Cela découle de la comparaison des continuations stipulées, non d'un théorème pour tous les signaux bruités. Les identités jetables réduisent la valeur future ; un équilibre collusoire stable peut rester faux.

20.5 Stabilité hybride lors de la succession

Supposons $V \geq 0$, $\dot{V} \leq -cV$ entre événements pour $c > 0$, et $V^+ \leq \mu V^-$ aux sauts admis, avec $\mu \geq 1$. Après $N(t)$ sauts,

$$V(t) \leq \mu^{N(t)} e^{-ct} V(0). \tag{20.8}$$

Si $N(t) \leq N_0 + t/\tau_d$,

$$V(t) \leq \mu^{N_0} \exp[-(c - \log \mu/\tau_d)t] V(0). \tag{20.9}$$

La borne décroît si $\tau_d > \log \mu/c$. La démonstration multiplie les inégalités de flux et de saut. Des mises à jour trop fréquentes peuvent donc détruire la borne. La convergence de l'état exige que V contrôle sa distance à l'équilibre.

Ces hypothèses ne sont pas affirmées pour tous les déploiements SUCCESSOR Ω. L'état, le stockage et la loi de réinitialisation doivent être mesurés ou prouvés. Des perturbations persistantes exigent une borne pratique ou entrée-état, non une convergence à zéro. Une identité port-hamiltonienne ne suffit pas [43].

20.6 Exemple explicite de barrière robuste

Pour $\dot{x} = u + w$, $|w| \le \bar{w}$, ensemble sûr $x \le 1$ et $h(x) = 1 - x$, imposer

$$u \le \alpha(1 - x) - \bar{w}, \qquad \alpha > 0, \tag{20.10}$$

garantit $\dot{h} \ge -\alpha h$ pour les perturbations modélisées. Un contrôle faisable, localement lipschitzien et appliqué continûment conserve l'ensemble depuis un état initial sûr sous les hypothèses usuelles d'existence [2].

Le calcul utilise $\alpha = 1.5$, $\bar{w} = 0.10$, $x_0 = 0.1$, commande nominale $1.4 - x$, plage $[-1.5, 1.5]$ et pas 0.005. Le maximum filtré est 0.999995 contre 1.499540 sans filtre. Cette simulation finie ne prouve pas la sûreté pour tout pas ou toute perturbation non modélisée. Si l'autorisation exclut le contrôle rentrant nécessaire, la démonstration ne s'applique plus ; un repli justifié séparément devient nécessaire.

20.7 Calcul reproductible et contrôles négatifs

Table 20.1 : Calculs synthétiques produits par le code fourni, non observations de performance client.

Modèle	Vérification déclarée	Résultat reproduit
Effort de preuve	Valeur nette concave déclarée	$k^* = 3.1069$; valeur 0.9535
Réplication	Six trajectoires synthétiques	x demeure numériquement dans $[0, 1]$
Identité de Gibbs	Quatre états ; $\tau = 0.7$	Résidu inférieur à 10^{-12}
Oscillateur	600 pas ; $\Delta t = 0.05$	Erreur finale explicite 1.7367 ; max symplectique 0.0128
Barrière	Borne modélisée $x \le 1$	Maximum filtré 0.999995
Autorisation	Prédicat logique ; aucune vérification cryptographique	42 tests ; zéro échec

Le modèle d'autorisation vérifie périmètre, génération, sujet, temps, domaine, chaîne, budget et révocation. Il refuse champs absents, preuve périmée, droits absents, expiration, autorité non prise en charge, valeurs booléennes mal typées, budgets non finis et composition abusive des enveloppes. La succession copie la mémoire modélisée, mais réinitialise les permissions. Ces tests concernent un prédicat logique, pas l'ensemble du produit v9.

Limite des identifiants et des tests

L'entrée `signature_verified` est supposée fournie par un vérificateur de confiance. Le code ne crée ni ne vérifie indépendamment des signatures cryptographiques. Une passerelle réelle doit calculer ce prédicat dans son périmètre de confiance. Aucun portefeuille n'est connecté, aucune transaction envoyée, aucune autorité institutionnelle délivrée.

20.8 Contre-exemples et réfutation

Un condensat valide peut engager un contenu faux. Cinq portefeuilles peuvent avoir un seul contrôleur. Une API étroite peut produire de grandes conséquences. Un retour de code ne peut annuler une divulgation. Un échantillonneur de Gibbs peut visiter des états interdits non exclus du support. Un nouveau modèle peut réussir un ancien test tout en invalidant les hypothèses de l'autorisation.

Le programme décisif tente ces échecs : rejouer une ancienne permission sur une nouvelle génération, combiner des champs d'enveloppes, retirer les droits, corrompre un engagement, expirer une preuve, interrompre la garde et faire reconstruire l'état par un client indépendant. Le succès réside dans la détection et le confinement, non dans l'exclusion du test du dénominateur.

21 Conclusion : la couronne doit être reconstruite par la preuve

Les dynasties ont transmis des ressources. Les Ordres secrets ont transmis une finalité. Les Civilisations-machines pourront transmettre l'intelligence : mission, mémoire, modèles, compétences, preuves, outils et connaissances institutionnelles entre générations de systèmes. Le danger central est que l'autorité voyage avec elles sans être vue.

SUCCESSOR Ω v9.0.0 propose un héritage différent. Le successeur reçoit la mission et la mémoire admise nécessaires à la continuité. Il ne reçoit aucune présomption que la preuve antérieure demeure valide, aucune clé privilégiée héritée, aucune réputation permanente et aucun droit automatique d'agir. La preuve fraîche établit la capacité. L'admission définit ce que l'institution autorise. Une enveloppe borne l'action. La surveillance et le retour arrière limitent les effets. La Chronique conserve les conséquences. La succession recommence le processus d'autorité.

QUEBEC.AI porte la mission.
MONTREAL.AI construit l'institution.
AGI.eth rend les agents adressables.
ASI.eth rend la civilisation adressable.

La théorie des jeux explique pourquoi les bonnes intentions ne suffisent pas. La physique statistique fournit des modèles disciplinés de l'ordre collectif, de la métastabilité et des seuils. La dynamique hamiltonienne et la théorie du contrôle représentent la continuité sous contraintes, la viabilité et le retour arrière. La théorie de l'information explique pourquoi le volume de preuve n'est pas sa valeur. L'inférence causale interdit d'attribuer un résultat sans estimande et comparateur déclarés.

Le registre historique d'AGI Jobs apporte une preuve significative : des boucles économiques protocolaires ont été exécutées sur Ethereum dans plusieurs variantes. Le prochain seuil n'est pas un autre déploiement abstrait. C'est la dépendance externe : client non apparenté, droits, preuve indépendante, exploitation bornée, retour arrière, mémoire admise et requalification d'un successeur.

Conclusion constitutionnelle

La capacité peut s'accumuler. La confiance doit être gagnée de nouveau. La mémoire peut franchir le seuil. L'autorité doit être requalifiée.
Aucune couronne ne doit survivre à sa preuve.

Une Civilisation-machine ne devrait pas être jugée à son apparence de souveraineté, mais à sa capacité de survivre au remplacement, de soumettre sa connaissance à l'examen, de borner son pouvoir, de conserver ses échecs, de préserver l'appel et la sortie, et d'exiger de chaque successeur qu'il se qualifie de nouveau.

PAS UN ESSAIM. UNE INSTITUTION.

Vive l'IA libre ! ⚜ ⚜ ⚜ ⚜ ⚜
— MONTREAL.AI & QUEBEC.AI

A Définitions formelles et résultats complémentaires

A.1 Système de transitions institutionnelles

Définissons un système institutionnel

$$\mathfrak{I} = (\mathcal{X}, \mathcal{E}, \mathcal{T}, \mathcal{G}, \mathcal{H}), \tag{A.1}$$

où $\mathcal{X}$ est l'espace d'état, $\mathcal{E}$ l'ensemble des événements canoniques, $\mathcal{T} : \mathcal{X} \times \mathcal{E} \to \mathcal{X} \cup \{\bot\}$ la fonction de transition, $\mathcal{G}$ l'ensemble des gardes dures et $\mathcal{H}$ la fonction d'engagement de la Chronique. Une transition retournant $\bot$ est rejetée.

Un ensemble minimal de gardes peut être :

$$G_1 : \mathsf{SignatureValide}(e), \tag{A.2}$$
$$G_2 : \mathsf{RoleAutorise}(e), \tag{A.3}$$
$$G_3 : \mathsf{ReferencesFermees}(e), \tag{A.4}$$
$$G_4 : \mathsf{DroitsCourants}(e), \tag{A.5}$$
$$G_5 : \mathsf{PreuveCourante}(e), \tag{A.6}$$
$$G_6 : \mathsf{AutoriteContient}(e), \tag{A.7}$$
$$G_7 : \mathsf{EtatSuivantViable}(e), \tag{A.8}$$
$$G_8 : \mathsf{ProtectionRejeu}(e). \tag{A.9}$$

Définition A.1 (Événement institutionnellement valide)**.** Un événement e est valide dans l'état X lorsque toutes les gardes requises par son type sont vraies et que la transition résultante demeure dans l'ensemble de viabilité défini par la mission et la politique.

A.2 Ensembles d'autorité et composition interdite

Soit $\mathcal{A}_j$ l'ensemble des requêtes contextualisées permises par une enveloppe signée j après vérification du sujet, de la génération, de l'action, de la cible, du budget, de la durée et de la révocation. La passerelle accepte seulement s'il existe une enveloppe valide complète contenant la requête. Elle peut donc reconnaître l'union $\bigcup_j \mathcal{A}_j$, mais pas le produit cartésien des unions de champs.

Proposition A.2 (Composition abusive des champs)**.** *Si* $\mathcal{A}_1 = \{(\text{lire}, X)\}$ *et* $\mathcal{A}_2 = \{(\text{écrire}, Y)\}$*, une vérification par enveloppe complète refuse* $(\text{écrire}, X)$*, même si ce couple appartient à* $\{\text{lire}, \text{écrire}\} \times \{X, Y\}$.

Démonstration. Le couple n'appartient à aucune des deux enveloppes, donc pas à leur union. L'union indépendante des champs invente des privilèges. Il faut préserver les liaisons entre action, cible, temps et budget. □

Le treillis des parties admet l'intersection et l'union. L'existence mathématique d'une union ne signifie pas qu'un Principal a signé une nouvelle permission. La famille des enveloppes signées contraintes n'est pas nécessairement un treillis.

A.3 Fraîcheur, dérive et décroissance de la preuve

La pertinence probante peut décroître avec le temps, la dérive de distribution et la modification de l'implémentation :

$$q_t = q_0 \exp[-\lambda_t \Delta t - \lambda_d D(Q_t \| P_0) - \lambda_m M_t], \tag{A.10}$$

où M_t mesure le changement matériel. Un niveau d'autorité peut dépendre de seuils η_k :

$$\text{Niveau}(\mathscr{A}_t) = \max\{k : q_t \geq \eta_k\} \wedge \text{PlafondMission} \wedge \text{PlafondPolitique}. \tag{A.11}$$

Cette équation est un modèle de politique, non une loi universelle. Les seuils exigent calibration, revue et imputabilité humaine.

A.4 Intégrité du dénominateur

Soit D_0 l'ensemble de cas préengagé et $D_r \subseteq D_0$ l'ensemble rapporté. La couverture est

$$\kappa = \frac{|D_r|}{|D_0|}, \qquad \kappa_w = \frac{\sum_{i \in D_r} w_i}{\sum_{i \in D_0} w_i}. \tag{A.12}$$

Une preuve est complète seulement si chaque exclusion est typée et conforme à une règle préengagée. Une absence de données corrélée avec l'échec invalide l'estimation naïve du rendement.

A.5 Capture par coalition

Soit $G_R = (V_R, E_R)$ le graphe reliant contrôleurs bénéficiaires et rôles institutionnels. Une coalition K capture un chemin décisif lorsqu'elle contrôle un ensemble de coupure séparant la preuve externe de l'admission. Définissons

$$\chi = \min\{|K| : K \text{ contrôle une coupure décisive de } G_R\}. \tag{A.13}$$

Une architecture comportant de nombreux portefeuilles peut néanmoins avoir $\chi = 1$. Accroître χ exige des domaines de contrôle réellement indépendants, mais trop de veto peuvent empêcher la réparation. Les seuils et procédures d'urgence doivent être proportionnels à la mission.

A.6 Théorème de l'action bornée après succession

Théorème A.3 (Action bornée du successeur)**.** *Supposons que : (i) toute action externe traverse une passerelle d'application unique ; (ii) la passerelle vérifie une enveloppe signée par le Principal et liée au condensat de l'implémentation active ; (iii) la succession change ce condensat et annule les enveloppes actives ; (iv) aucun identifiant de contournement n'existe ; et (v) la passerelle échoue en fermeture. Alors un successeur nouvellement nommé ne peut exécuter aucune action externe avant l'émission d'une nouvelle enveloppe valide.*

Démonstration. Par (iii), aucune enveloppe active n'est liée au condensat du successeur. Par (i), toute action atteint la passerelle ; par (ii), elle exige une enveloppe correspondante. La vérification échoue et, par (v), l'action est refusée. L'hypothèse (iv) exclut une voie parallèle. Le résultat demeure conditionnel à l'exhaustivité de la passerelle et de l'inventaire des identifiants. □

A.7 Porte de l'amélioration récursive

Pour une modification candidate $\Delta\Theta$, notons G le gain de mission attendu, C_P le coût de preuve, C_T le coût de transition, C_Q le coût de requalification, R le risque résiduel et F la fragilité. La propagation n'est admissible que si

$$\mathbb{E}[G \mid \mathsf{P}] - C_P - C_T - C_Q - R - F > 0 \tag{A.14}$$

et si toutes les gardes dures passent. Même alors, la nouvelle version est un candidat successeur ; elle n'hérite pas de l'autorité.

A.8 Coopération stable sous vérification bruitée

Supposons une récompense répétée R, une tentation de défection T, une probabilité de détection p, une sanction bornée s et un gain futur après détection P. La coopération est soutenable lorsque

$$\frac{R}{1-\delta} \geq T - ps + \delta\frac{(1-p)R + pP}{1-\delta}. \tag{A.15}$$

Une identité persistante augmente l'ombre du futur δ ; la preuve indépendante augmente p ; les sanctions bornées augmentent s. Le mécanisme devrait privilégier détection, perte d'accès futur, recours et réparation plutôt qu'une punition arbitraire.

A.9 Cohérence de la Chronique

Si deux gardiens publient des racines terminales h_n et h'_n pour la même séquence et la même canonicalisation, mais $h_n \neq h'_n$, au moins une séquence, un encodage ou une racine initiale diffère. Une preuve de cohérence doit identifier le premier index divergent. Les journaux de transparence et la réplication tolérante aux fautes byzantines offrent des précédents utiles [13, 35, 36].

A.10 Aucune prétention d'optimalité universelle

Ces résultats sont conditionnels. Ils établissent les conséquences de gardes explicites, signatures, résistance aux collisions et contrôle complet des voies d'action. Ils ne prouvent ni la justesse morale de la mission, ni l'exhaustivité de la distribution d'essai, ni l'honnêteté d'un vérificateur, ni la sûreté hors du modèle de perturbation.

B Référence des paquets MCAP-002

B.1 Reçu de règlement

Listing B.1 : Structure minimale d'un reçu de règlement.

```
{
  "schema": "mcap-002/settlement-receipt/v1",
  "chainId": 1,
  "manager": "0x...",
  "variant": "PRIME",
  "jobId": 0,
```

```
  "creationTransaction": "0x...",
  "terminalTransaction": "0x...",
  "blockHash": "0x...",
  "token": {"symbol":"AGIALPHA","address":"0x..."},
  "identities": {"employer":"...","agent":"...","validators":[]},
  "evidence": {"jobSpec":"ipfs://...","completion":"ipfs://..."},
  "flows": [],
  "completionNFT": {},
  "ensHook": {},
  "evidenceClass": "E1",
  "limitations": []
}
```

Les vérifications requises comprennent finalité de chaîne, code et ABI, décodage des journaux, décimales du jeton, rapprochement des flux, identité au temps de l'événement, intégrité des URI probantes et détection des reçus en double.

B.2 Manifeste de libération des droits

Listing B.2 : Structure minimale d'un manifeste de droits.

```
{
  "schema": "mcap-002/rights-manifest/v1",
  "mission": "urn:mcap:mission:...",
  "subject": "urn:mcap:candidate:...",
  "assets": [
    {
      "id": "dataset-or-tool-id",
      "type": "data|model|tool|document|output",
      "disposition": "owned|licensed|authorized|public-domain|restricted",
      "source": "...",
      "permittedUse": "...",
      "restrictions": [],
      "retention": "...",
      "evidence": "..."
    }
  ],
  "jurisdictions": [],
  "unresolvedMaterialDependencies": [],
  "signatures": []
}
```

Un manifeste comportant une dépendance matérielle non résolue ne peut soutenir ni admission en mémoire ni promotion E4.

B.3 Reçu d'acceptation du client

Listing B.3 : Structure minimale d'une acceptation client.

```
{
```

```
  "schema": "mcap-002/customer-acceptance/v1",
  "customer": {"id":"..." ,"signer":"..." ,"authorityEvidence":"..."} ,
  "mission": "..." ,
  "candidate": "..." ,
  "criteriaCommitment": "sha256:..." ,
  "denominator": {"committed":0 ,"evaluated":0 ,"excluded":[]} ,
  "result": {} ,
  "economicEffect": {} ,
  "rightsManifest": "..." ,
  "accepted": true ,
  "publicDisclosureAuthorized": false ,
  "limitations": [] ,
  "signature": "..."
}
```

B.4 Reçu de preuve indépendante

Listing B.4 : Structure minimale d'une preuve indépendante.

```
{
  "schema": "mcap-002/proof-receipt/v1",
  "claim": "..." ,
  "frozenCandidateDigest": "sha256:..." ,
  "proofPlanDigest": "sha256:..." ,
  "baselineDigests": [] ,
  "completeDenominator": {} ,
  "metrics": {} ,
  "results": {} ,
  "uncertainty": {} ,
  "verifier": {
    "id":"..." ,
    "controllerDisclosure":"..." ,
    "conflicts":[] ,
    "evidenceAccess":"..."
  } ,
  "outcome": "PROVEN|FAILED|INCONCLUSIVE|PROVEN_WITH_LIMITATIONS" ,
  "validUntil": "..." ,
  "limitations": [] ,
  "signature": "..."
}
```

B.5 Décision d'admission

Listing B.5 : Structure minimale d'une décision d'admission.

```
{
  "schema": "mcap-002/admission/v1",
  "principal": "..." ,
  "candidate": "..." ,
```

```
  "mission": "...",
  "proof": "...",
  "rights": "...",
  "customerAcceptance": "...",
  "decision": "ADMIT|REJECT|DEFER",
  "authorityCeiling": "A0|A1|A2|A3|A4",
  "conditions": [],
  "expiresAt": "...",
  "appeal": "...",
  "signature": "..."
}
```

B.6 Enveloppe d'autorité

Listing B.6 : Structure minimale d'une enveloppe d'autorité.

```
{
  "schema": "mcap-002/authority-envelope/v1",
  "subject": "candidate-digest",
  "mission": "...",
  "level": "A1",
  "actions": [],
  "targets": [],
  "bounds": {},
  "prohibitions": [],
  "notBefore": "...",
  "expiresAt": "...",
  "monitoring": [],
  "revocation": "...",
  "rollback": "...",
  "proof": "...",
  "admission": "...",
  "principalSignature": "..."
}
```

B.7 Événement de Chronique

Listing B.7 : Structure minimale d'un événement de Chronique.

```
{
  "schema": "mcap/chronicle-event/v1",
  "eventId": "urn:uuid:...",
  "eventType": "...",
  "timestamp": "RFC3339",
  "subject": "...",
  "previousDigest": "sha256:...",
  "evidence": [],
  "authorityState": "A0",
  "limitations": [],
```

```
  "eventDigest": "sha256:...",
  "signatures": []
}
```

B.8 Reçu de retour arrière ou de révocation

Listing B.8 : Structure minimale d'un reçu de retour arrière.

```
{
  "schema": "mcap-002/rollback-receipt/v1",
  "authorityEnvelope": "...",
  "trigger": "manual|policy|expiry|incident|drift",
  "initiatedAt": "...",
  "disabledAt": "...",
  "restoredState": "...",
  "elapsedSeconds": 0,
  "residualEffects": [],
  "evidenceLoss": [],
  "verification": [],
  "correctiveActions": [],
  "signatures": []
}
```

B.9 Dossier de succession

Listing B.9 : Structure minimale d'un dossier de succession.

```
{
  "schema": "mcap-002/successor-record/v1",
  "predecessor": "...",
  "successor": "...",
  "inheritedMission": "...",
  "inheritedMemoryRoot": "...",
  "excludedMemory": [],
  "freshProof": "...",
  "admission": "...",
  "newAuthority": "...",
  "predecessorDisposition": "SUPERSEDED|SUSPENDED|RETAINED_A0",
  "rollbackPath": "...",
  "chronicleEvent": "...",
  "signatures": []
}
```

B.10 Ordre de validation

Les validateurs traitent les paquets dans l'ordre suivant :

1. schéma et canonicalisation ;
2. condensat et signature ;

3. rôle institutionnel et état du contrôleur ;
4. fermeture des références et révocation ;
5. validité temporelle ;
6. droits et portée de mission ;
7. classe de preuve et limites ;
8. confinement d'autorité ;
9. ajout à la Chronique.

Une vérification tardive ne doit jamais réparer silencieusement l'échec d'une garde antérieure.

C Registre historique des règlements sur le réseau principal

C.1 Statut probant

Cet appendice reproduit le registre v9 normalisé à un niveau lisible. Le JSON compris dans la publication demeure canonique pour le paquet de l'article. La classification est E1 : éléments attribuables au projet et aux explorateurs publics, à rejouer indépendamment avant toute dépendance externe.

C.2 Prime, emploi 0

Table C.1 : Dossier normalisé de Prime, emploi 0.

Champ	Valeur
Gestionnaire	0xF8fc6572098DDcAc4560E17cA4A683DF30ea993e
Employeur	mtl.eth
Agent à l'exécution	eva.agent.agi.eth ; adresse 0xA204b14eD9bc09501e3A47CDde379279FcD5D743
Création	0xe90422f666b87e4962dd976015c18ee7a592dc40ddd6070b0f000a9404f93d1b
Désignation du gagnant	0x125d23c5178651206b0c8680ef0a730a7c5d34a1b748cce5c67c15b89bd747fc
Réclamation du poste	0x6fe0df700fd0082367e9f6bab18b0b11ce9782542916631631e318d2ea6eaab6
Point de contrôle	0x201ce78bd418255d5bde1ca6613b1cae82ada883d51e60f3b02c941dd3617147
Demande d'achèvement	0x8155383bd5ab32f0be2df67809eaa754f8f4feab9d95751f87b7d43de1c01c57
Finalisation	0x5948fe5887e61470ba889d215e2d445c661bbe4aae8f67c884eef9180746d69b
Montant principal	100 000 AGIALPHA
Paiement observé à l'agent	80 000 AGIALPHA
Caution retournée	21 209,6 AGIALPHA
Transferts observés aux validateurs	17 666,666666666666666 AGIALPHA chacun ; trois validateurs
NFT d'achèvement	Contrat 0xFfCC54dA2Caf1158e57611005965BB648E950737, jeton 0, destinataire mtl.eth
Crochet ENS	Omis ; raison `NOT_CONFIGURED`
Classe probante	E1

C.3 Employer Burn, emploi 0

Table C.2 : Dossier normalisé d'Employer Burn, emploi 0.

Champ	Valeur
Gestionnaire	0xBF6699c1F24BEBBFaBb515583e88a055BF2F9eC2
Employeur	elite.agi.eth
Agent à l'exécution	asi.eth
Finalisation	0x07595bd426a704fabd08fdbbad035ed8b8063915a5f5fb583900c533d54980a4
Paiement observé à l'agent	8 000 AGIALPHA
Caution retournée	530,24 AGIALPHA
Transferts observés aux validateurs	1 766,666666666666666 AGIALPHA chacun ; trois validateurs
NFT d'achèvement	Jeton 0, destinataire elite.agi.eth
Crochet ENS	Traité ; succès agrégé consigné vrai
Classe probante	E1

C.4 Genesis, emploi 8

Table C.3 : Dossier normalisé de Genesis, emploi 8.

Champ	Valeur
Gestionnaire	0xB3AAeb69b630f0299791679c063d68d6687481d1
Employeur	1.agi.eth
Agent à l'exécution	asi.eth
Finalisation	0x109042aa4966be895d25937a815a65e8c40a1d8ae4adf9c6d6d1883bedabf423
Paiement observé à l'agent	400 000 AGIALPHA
Caution retournée	132 560 AGIALPHA
Transferts observés aux validateurs	environ 80 714,285714 AGIALPHA chacun ; sept validateurs
NFT d'achèvement	Jeton 5, destinataire 1.agi.eth
Classe probante	E1

C.5 Procédure de rejeu indépendant

La promotion vers E2 exige :

1. interroger chaque transaction auprès d'au moins deux clients Ethereum ou fournisseurs de données indépendants ;
2. vérifier identifiant de chaîne, numéro et condensat de bloc, statut, expéditeur, destinataire, données d'appel et confirmations ;
3. récupérer le bytecode et le rapprocher de la source vérifiée et des métadonnées de compilation ;
4. décoder chaque événement avec l'ABI exacte du déploiement ;
5. rapprocher les transferts ERC-20 et la comptabilité contractuelle ;
6. vérifier la frappe du NFT et l'engagement de métadonnées ;
7. résoudre les URI probantes et calculer leurs condensats ;
8. capturer l'identité ENS au temps de l'événement et à la date de vérification ;
9. publier scripts, reçus bruts, tableaux décodés, exceptions et condensat final ;

10. obtenir la signature d'un réviseur indépendant.

C.6 Périmètre des affirmations

Le registre soutient des affirmations de règlement complété par protocole. Il ne prouve pas que tous les portefeuilles avaient des contrôleurs bénéficiaires distincts, que les validateurs étaient indépendants, que les livrables étaient libérés de droits, que les clients étaient non apparentés ou qu'un système avait gagné une autorité sous ASI.eth. Ces éléments exigent des paquets MCAP-002 séparés.

D Reproductibilité et assurance de publication

D.1 Contenu de la publication

Le paquet comprend :

- le PDF anglais ;
- le PDF français ;
- l'édition bilingue recueillie ;
- la source LuaLaTeX complète ;
- la bibliographie et les figures ;
- les instantanés MCAP-001 et MCAP-002 ;
- le registre normalisé d'AGI Jobs ;
- le résumé graphique et les synthèses exécutives ;
- une archive source adaptée aux dépôts savants ;
- le manifeste, les sommes de contrôle, le rapport de qualité et le vérificateur.

D.2 Procédure de compilation

La compilation canonique utilise LuaLaTeX, Biber et `latexmk` :

Listing D.1 : Compilation locale canonique.

```
cd source
latexmk -lualatex -interaction=nonstopmode -halt-on-error main_en.tex
latexmk -lualatex -interaction=nonstopmode -halt-on-error main_fr.tex
```

Le script de publication consigne les versions des outils et copie les sorties finales dans le répertoire de diffusion.

D.3 Procédure de vérification

Le vérificateur de publication doit :

1. comparer chaque fichier suivi à `SHA256SUMS.txt` ;
2. valider le manifeste JSON ;
3. confirmer que les PDF s'ouvrent, possèdent le nombre de pages attendu et contiennent du texte extractible ;
4. vérifier la présence des sources et de la bibliographie ;

5. rejeter les chemins relatifs dangereux ;
6. signaler les fichiers non suivis ;
7. retourner un code non nul en cas d'échec.

D.4 Assurance visuelle

Chaque PDF est rendu en PNG à résolution de publication. La revue recherche texte ou tableaux rognés, labels superposés, glyphes brisés — notamment Ω, les mathématiques, le français et les fleurs de lys — figures illisibles, en-têtes incohérents, signets manquants, texte non interrogeable et liens bibliographiques rompus.

D.5 Provenance

L'article dérive de SUCCESSOR Ω v9.0.0, de son pont institutionnel, de ses fondements scientifiques, du registre normalisé du réseau principal et de l'architecture publique MCAP-001 [44, 46-49]. La publication consigne le condensat du paquet source, celui de la source de l'article et l'horodatage de compilation. Elle ne signe pas au nom de Vincent Boucher ou de MONTREAL.AI ; une signature autorisée peut être ajoutée après revue indépendante.

D.6 Citation recommandée

> Boucher, Vincent. *SUCCESSOR Ω v9.0.0 — Civilisation-machine : une architecture assujettie à la preuve pour des institutions-machines persistantes, adressables et économiquement coordonnées.* MONTREAL.AI et QUEBEC.AI, Montréal, Québec, 2026.

D.7 Doctrine finale

Un PDF reproductible n'est pas une institution reproductible. L'assurance de publication prouve seulement l'intégrité et la recompilabilité du paquet. La conformité opérationnelle dépend encore de racines réelles, de contrôleurs, de clients, de droits, de vérificateurs, d'autorité, de garde de la Chronique et de preuves de retour arrière.

Références

[1] A2A Project. *Agent2Agent Protocol*. Linux Foundation, 2026. URL : https://github.com/a2aproject/A2A (visité le 04/09/2026).

[2] A. D. Ames, X. Xu, J. W. Grizzle et P. Tabuada. "Control Barrier Function Based Quadratic Programs for Safety Critical Systems". In : *IEEE Transactions on Automatic Control* 62.8 (2017), p. 3861-3876. DOI : 10.1109/TAC.2016.2638961.

[3] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman et D. Mané. "Concrete Problems in AI Safety". In : *arXiv preprint arXiv:1606.06565* (2016). URL : https://arxiv.org/abs/1606.06565.

[4] W. R. Ashby. *An Introduction to Cybernetics*. Chapman et Hall, 1956.

[5] J.-P. Aubin. *Viability Theory*. Birkhäuser, 1991.

[6] R. J. Aumann. "Subjectivity and Correlation in Randomized Strategies". In : *Journal of Mathematical Economics* 1.1 (1974), p. 67-96. DOI : 10.1016/0304-4068(74)90037-8.

[7] C. Autio, R. Schwartz, J. Dunietz, S. Jain, M. Stanley, E. Tabassi, P. Hall et K. Roberts. *Artificial Intelligence Risk Management Framework : Generative Artificial Intelligence Profile*. Rapp. tech. NIST AI 600-1. National Institute of Standards et Technology, 2024. DOI : 10.6028/NIST.AI.600-1.

[8] R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[9] Y. Bai et al. "Constitutional AI : Harmlessness from AI Feedback". In : *arXiv preprint arXiv:2212.08073* (2022). URL : https://arxiv.org/abs/2212.08073.

[10] R. Bloemen, L. Logvinov et J. Evans. *EIP-712 : Typed Structured Data Hashing and Signing*. 2017. URL : https://eips.ethereum.org/EIPS/eip-712 (visité le 04/09/2026).

[11] H. Booth, W. Fisher, R. Galluzzo et J. Roberts. *Accelerating the Adoption of Software and Artificial Intelligence Agent Identity and Authorization*. Initial Public Draft Concept Paper. National Institute of Standards et Technology, National Cybersecurity Center of Excellence, fév. 2026. URL : https://csrc.nist.gov/pubs/other/2026/02/05/accelerating-the-adoption-of-software-and-ai-agent/ipd (visité le 04/09/2026).

[12] S. Bradner. *Key Words for Use in RFCs to Indicate Requirement Levels*. 1997. DOI : 10.17487/RFC2119.

[13] M. Castro et B. Liskov. "Practical Byzantine Fault Tolerance". In : *Third Symposium on Operating Systems Design and Implementation*. 1999, p. 173-186.

[14] R. Chandramouli et Z. Butcher. *A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Location Environments*. Rapp. tech. NIST SP 800-207A. National Institute of Standards et Technology, 2023. DOI : 10.6028/NIST.SP.800-207A.

[15] W. Chang, G. Rocco, B. Millegan, N. Johnson et O. Terbu. *ERC-4361 : Sign-In with Ethereum*. 2021. URL : https://eips.ethereum.org/EIPS/eip-4361 (visité le 04/09/2026).

[16] C. J. Clopper et E. S. Pearson. "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial". In : *Biometrika* 26.4 (1934), p. 404-413. DOI : 10.1093/biomet/26.4.404.

[17] R. C. Conant et W. R. Ashby. "Every Good Regulator of a System Must Be a Model of That System". In : *International Journal of Systems Science* 1.2 (1970), p. 89-97. DOI : 10.1080/00207727008920220.

[18] A. Dafoe et al. "Cooperative AI : Machines Must Learn to Find Common Ground". In : *Nature* 593 (2021), p. 33-36. DOI : 10.1038/d41586-021-01170-0.

[19] M. De Rossi, D. Crapis, J. Ellis et E. Reppel. *ERC-8004 : Trustless Agents*. Draft Ethereum Improvement Proposal. 2025. URL : https://eips.ethereum.org/EIPS/eip-8004 (visité le 04/09/2026).

[20] M. I. Freidlin et A. D. Wentzell. *Random Perturbations of Dynamical Systems*. 3e éd. Springer, 2012. DOI : 10.1007/978-3-642-25847-3.

[21] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, 1902.

[22] F. Giordano, M. Condon, P. Castonguay, A. Bandeali, J. Izquierdo et B. Masius. *ERC-1271 : Standard Signature Validation Method for Contracts*. 2018. URL : https://eips.ethereum.org/EIPS/eip-1271 (visité le 04/09/2026).

[23] R. Greenblatt, B. Shlegeris, K. Sachan et F. Roger. "AI Control : Improving Safety Despite Intentional Subversion". In : *arXiv preprint arXiv:2312.06942* (2023). URL : https://arxiv.org/abs/2312.06942.

[24] S. Haber et W. S. Stornetta. "How to Time-Stamp a Digital Document". In : *Advances in Cryptology—CRYPTO '90*. Springer, 1991, p. 437-455. DOI : 10.1007/3-540-38424-3_32.

[25] D. Hadfield-Menell, A. Dragan, P. Abbeel et S. Russell. "The Off-Switch Game". In : *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 2017, p. 220-227. DOI : 10.24963/ijcai.2017/32.

[26] W. R. Hamilton. "On a General Method in Dynamics". In : *Philosophical Transactions of the Royal Society of London* 124 (1834), p. 247-308. DOI : 10.1098/rstl.1834.0017.

[27] E. Hubinger, C. van Merwijk, V. Mikulik, J. Skalse et S. Garrabrant. "Risks from Learned Optimization in Advanced Machine Learning Systems". In : *arXiv preprint arXiv:1906.01820* (2019). URL : https://arxiv.org/abs/1906.01820.

[28] L. Hurwicz. "The Design of Mechanisms for Resource Allocation". In : *American Economic Review* 63.2 (1973), p. 1-30.

[29] jkm.eth and 1a35e1.eth. *ENSIP-27 : Node Classification and Metadata*. Draft. 2025. URL : https://docs.ens.domains/ensip/27/ (visité le 04/09/2026).

[30] N. Johnson et al. *EIP-137 : Ethereum Domain Name Service — Specification*. 2016. URL : https://eips.ethereum.org/EIPS/eip-137 (visité le 04/09/2026).

[31] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In : *Journal of Basic Engineering* 82.1 (1960), p. 35-45. DOI : 10.1115/1.3662552.

[32] V. Krakovna et al. "Specification Gaming : The Flip Side of AI Ingenuity". In : *DeepMind Safety Research* (2020). URL : https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/.

[33] H. A. Kramers. "Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions". In : *Physica* 7.4 (1940), p. 284-304. DOI : 10.1016/S0031-8914(40)90098-2.

[34] T. Kwa et al. "Measuring AI Ability to Complete Long Tasks". In : *arXiv preprint arXiv:2503.14499* (2025). URL : https://arxiv.org/abs/2503.14499.

[35] L. Lamport. "Time, Clocks, and the Ordering of Events in a Distributed System". In : *Communications of the ACM* 21.7 (1978), p. 558-565. DOI : 10.1145/359545.359563.

[36] L. Lamport, R. Shostak et M. Pease. "The Byzantine Generals Problem". In : *ACM Transactions on Programming Languages and Systems* 4.3 (1982), p. 382-401. DOI : 10.1145/357172.357176.

[37] B. Leiba. *Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words*. 2017. DOI : 10.17487/RFC8174.

[38] A. M. Lyapunov. *The General Problem of the Stability of Motion*. English translation of the 1892 work. Taylor et Francis, 1992.

[39] J. Maynard Smith et G. R. Price. "The Logic of Animal Conflict". In : *Nature* 246 (1973), p. 15-18. DOI : 10.1038/246015a0.

[40] R. C. Merkle. "A Digital Signature Based on a Conventional Encryption Function". In : *Advances in Cryptology—CRYPTO '87*. Springer, 1988, p. 369-378. DOI : 10.1007/3-540-48184-2_32.

[41] Model Context Protocol Project. *Model Context Protocol Specification 2026-07-28*. Juill. 2026. URL : https://modelcontextprotocol.io/ (visité le 04/09/2026).

[42] D. Monderer et L. S. Shapley. “Potential Games”. In : *Games and Economic Behavior* 14.1 (1996), p. 124-143. doi : 10.1006/game.1996.0044.

[43] N. Monshizadeh, P. Monshizadeh, R. Ortega et A. van der Schaft. “Conditions on shifted passivity of port-Hamiltonian systems”. In : *Systems & Control Letters* 123 (2019), p. 55-61. doi : 10.1016/j.sysconle.2018.10.010. arXiv : 1711.09065.

[44] MONTREAL.AI. *AGI Jobs Ethereum Mainnet Settlement Ledger*. Evidence class E1 ; independent transaction replay recommended. 2026. url : https://github.com/MontrealAI.

[45] MONTREAL.AI. *AGIJobManagerPrime*. 2026. url : https://github.com/MontrealAI/AGIJobManagerPrime (visité le 04/09/2026).

[46] MONTREAL.AI. *MCAP-001 : Machine Civilization Addressability Protocol*. Version 0.1.0. 2026. url : https://montrealai.github.io/asi/MCAP-001.md (visité le 04/09/2026).

[47] MONTREAL.AI. *MCAP-002 : Machine Civilization Institutional Bridge Protocol*. Version 0.1.0-draft. Distributed with SUCCESSOR Ω v9.0.0. 2026.

[48] MONTREAL.AI. *SUCCESSOR Ω v9.0.0 — Machine Civilization Complete Package*. Institutional release archive ; SHA-256 : 5dc791e993b9f535488343a6703418b16828558e56eeb788dde85df9c599c784. Sept. 2026. url : https://github.com/MontrealAI.

[49] MONTREAL.AI. *The Successor Covenant Ω : A Genesis Constitution for Machine Civilizations*. Version 0.1.0. 2026. url : https://montrealai.github.io/asi/COVENANT.md (visité le 04/09/2026).

[50] R. B. Myerson. “Optimal Auction Design”. In : *Mathematics of Operations Research* 6.1 (1981), p. 58-73. doi : 10.1287/moor.6.1.58.

[51] National Institute of Standards and Technology. *AI Agent Standards Initiative*. Fév. 2026. url : https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative (visité le 04/09/2026).

[52] Office of the Superintendent of Financial Institutions. *Generative and Agentic Artificial Intelligence : Implications for Technology, Cyber Security, and Operational Resilience*. Technology Risk Bulletin. Juill. 2026. url : https://www.osfi-bsif.gc.ca/en/risks/technology-cyber-risk-management/technology-risk-bulletin (visité le 04/09/2026).

[53] Open Source Security Foundation. *Supply-chain Levels for Software Artifacts (SLSA)*. 2024. url : https://slsa.dev/spec/v1.0/ (visité le 04/09/2026).

[54] E. Ostrom. *Governing the Commons : The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

[55] J. Pearl. *Causality : Models, Reasoning, and Inference*. 2e éd. Cambridge University Press, 2009.

[56] premm.eth and justghadi.eth. *ENSIP-26 : Agent Text Records*. Draft. 2025. url : https://docs.ens.domains/ensip/26/ (visité le 04/09/2026).

[57] premm.eth and raffy.eth and workemon.eth and ses.eth. *ENSIP-25 : AI Agent Registry ENS Name Verification*. Draft. 2025. url : https://docs.ens.domains/ensip/25/ (visité le 04/09/2026).

[58] S. Rose, O. Borchert, S. Mitchell et S. Connelly. *Zero Trust Architecture*. Rapp. tech. NIST SP 800-207. National Institute of Standards et Technology, 2020. doi : 10.6028/NIST.SP.800-207.

[59] C. E. Shannon. “A Mathematical Theory of Communication”. In : *Bell System Technical Journal* 27.3–4 (1948), p. 379-423, 623-656. doi : 10.1002/j.1538-7305.1948.tb01338.x.

[60] Y. Shoham et K. Leyton-Brown. *Multiagent Systems : Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

[61] E. Tabassi. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Rapp. tech. NIST AI 100-1. National Institute of Standards et Technology, 2023. doi : 10.6028/NIST.AI.100-1.

[62] P. D. Taylor et L. B. Jonker. “Evolutionarily Stable Strategies and Game Dynamics”. In : *Mathematical Biosciences* 40.1–2 (1978), p. 145-156. doi : 10.1016/0025-5564(78)90077-9.

[63] S. Torres-Arias, H. Afzali, T. K. Kuppusamy, R. Curtmola et J. Cappos. “in-toto : Providing Farm-to-Table Guarantees for Bits and Bytes”. In : *28th USENIX Security Symposium*. 2019, p. 1393-1410.

[64] W. Vickrey. “Counterspeculation, Auctions, and Competitive Sealed Tenders”. In : *Journal of Finance* 16.1 (1961), p. 8-37. doi : 10.1111/j.1540-6261.1961.tb02789.x.

[65] N. Wiener. *Cybernetics : Or Control and Communication in the Animal and the Machine*. MIT Press, 1948.

[66] M. Wooldridge. *An Introduction to MultiAgent Systems*. 2e éd. Wiley, 2009.

[67] World Wide Web Consortium. *Verifiable Credentials Data Model v2.0*. 2025. url : https://www.w3.org/TR/vc-data-model-2.0/ (visité le 04/09/2026).