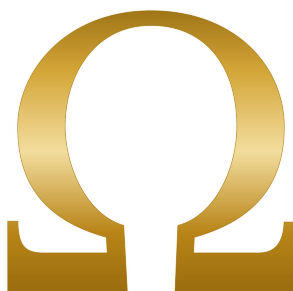


Proof-Gated Frontier Mission Intelligence

Strategic Equilibria · Information Thermodynamics · Statistical Physics · Hamiltonian Mission Control · Gibbs-Style Institutional Free Energy

Intelligence de mission de pointe assujettie à la preuve

Équilibres stratégiques · Thermodynamique de l'information · Physique statistique · Contrôle hamiltonien de mission · Énergie libre institutionnelle de style Gibbs



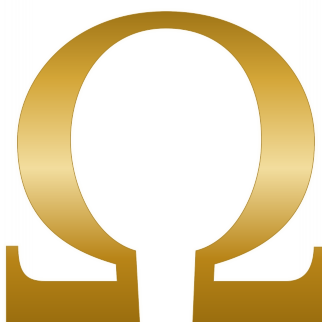
$$qF \geq b \quad | \quad \pi_{\tau}(\omega) \propto e^{-H/\tau} \quad | \quad \mathcal{H}_P = \ell + \lambda^{\top} f + B_A \quad | \quad C \neq P \neq D \neq A$$

EN · Scope. Artifact-grounded frontier simulation. Establishes a customer-controlled, mission-trained Specialist candidate, local Alpha*, complete-denominator underwriting, adversarial stress and a verifier-ready proof plan under A0. It does not establish real-world Alpha, independent proof, production admission or higher authority.

FR · Portée. Simulation de pointe fondée sur les artéfacts. Elle établit un candidat spécialiste entraîné pour la mission et sous le contrôle du client, un Alpha* local, une souscription à dénominateur complet, un stress contradictoire et un plan de preuve prêt pour vérification sous A0. Elle n'établit ni Alpha réel, ni preuve indépendante, ni admission en production, ni autorité supérieure.

Proof-Gated Frontier Mission Intelligence

Strategic Equilibria · Information Thermodynamics · Statistical Physics
Hamiltonian Mission Control · Gibbs-Style Institutional Free Energy



A complete artifact-grounded institutional grand synthesis connecting exact neural-symbolic candidate formation, comparator-relative Mission Alpha, portfolio underwriting, information gain, evolutionary verification games, Hamilton-Jacobi reachability, large-deviation refusal boundaries, and constrained Gibbs free-energy governance—while preserving independent proof, accountable admission, and effective authority as separate institutional decisions.

GAME THEORY

Truthful equilibria · role separation

STATISTICAL PHYSICS

Admissible ensembles · uncertainty

HAMILTONIAN CONTROL

Reversible search · safe reachability

GIBBS FREE ENERGY

Value · risk · cost · admissibility

$$qF \geq b \quad | \quad \pi_\tau(\omega) \propto e^{-\mathcal{H}/\tau} \quad | \quad \mathcal{H}_P = \ell + \lambda^\top f + B_A \quad | \quad C \neq P \neq D \neq A$$

FRONTIER PROOF CANDIDATE

A0 · 0 EXTERNAL EFFECTS

ARTIFACT-GROUNDED

Vincent Boucher · President, MONTREAL.AI & QUEBEC.AI

Grand Institutional Monograph v8.0.0
2026-09-04

Institutional Decision Brief

DECISION IN ONE SENTENCE

Proceed to a 90-day independently governed A0 shadow proof of the exact frozen candidate; do not grant production admission, consequential authority, realized-value credit, or the Specialist ASI designation until fresh evidence satisfies the precommitted proof contract.

CA\$ 48.8B fictional annual supplier-payment flow	1,000,000 weighted annual decision opportunities	CA\$ 218.5M central net scenario Alpha after complete deductions
CA\$ 70.5M conservative lower scenario after complete deductions	98.27% simulated $P(\text{net Alpha} > \text{CA\\$ } 50\text{M})$	A0 / 0 effective authority / external effects

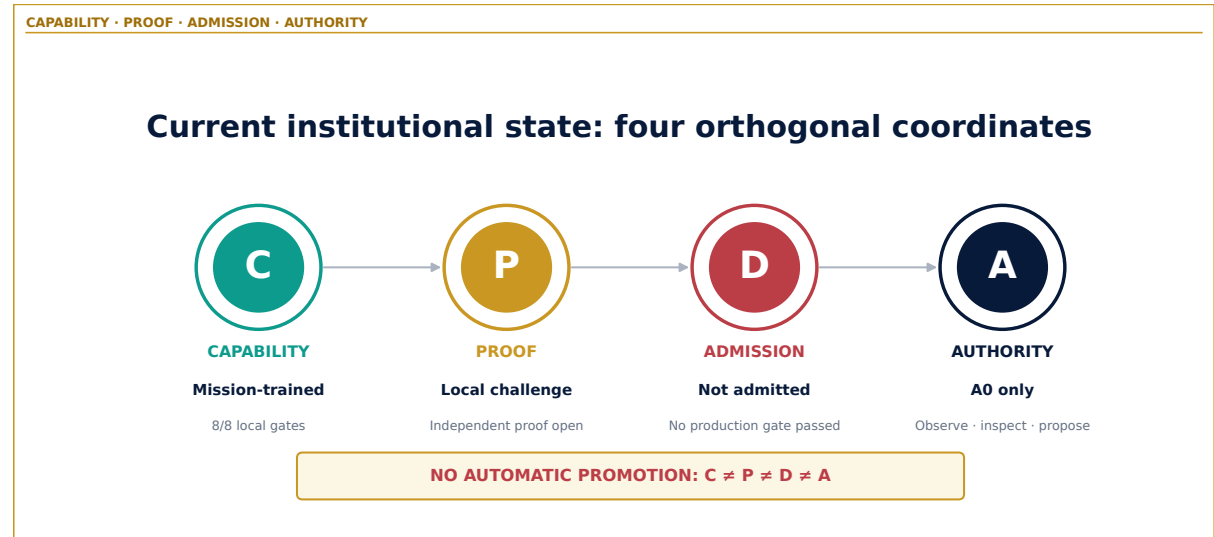
Board-level interpretation

The candidate is strong enough to justify an expensive independent proof programme, but not strong enough to justify skipping it. The local signal survives explicit transfer loss, complete-denominator deductions, distributional uncertainty, causal-transfer sensitivity, and multiple adversarial ecologies. The study does not establish real-world Alpha, independent proof, production fitness, or a right to act.

Table 1.: Decision rights and current evidentiary state.

Coordinate	Current state	Decision
Capability C	Useful, mission-trained, frozen neural-symbolic Specialist candidate	Continue evaluation
Proof P	Local synthetic holdout passed; no independent fresh-evidence proof	Commission independent proof
Admission D	No customer-production admission	Withhold
Authority A	A0: observe, inspect, and propose only; zero external effects	Maintain A0

Institutional State Architecture



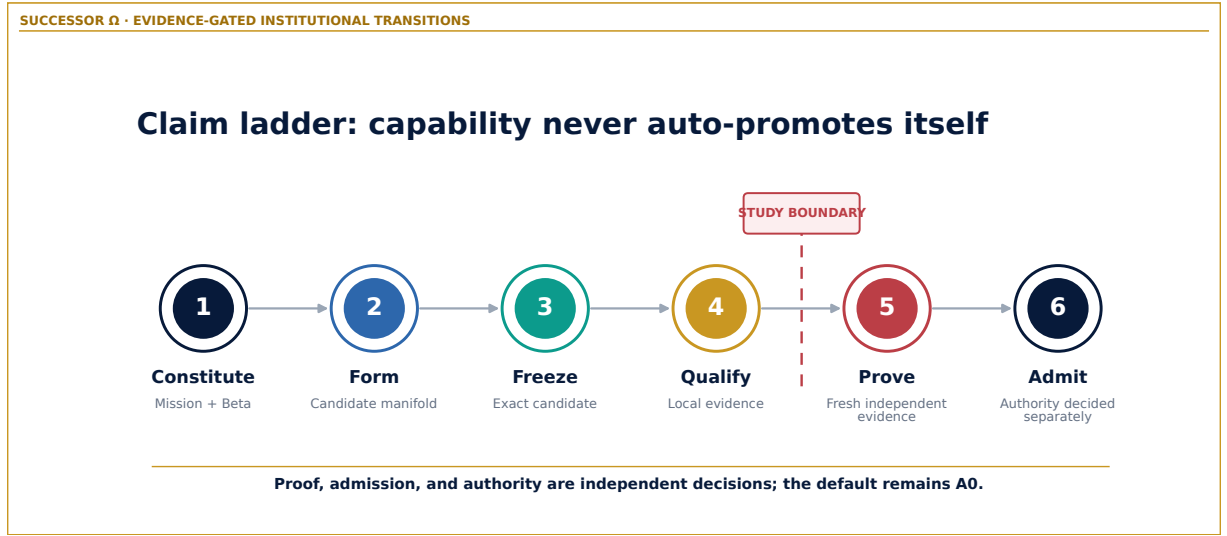
Capability, proof, admission, and authority are four orthogonal institutional coordinates.

The governing invariant

Capability ≠ Proof ≠ Admission ≠ Authority

A successful proof changes only the proof coordinate. Admission and authority remain unchanged until separately constituted decisions are recorded.

Evidence Ladder and Refusal Bound-ary

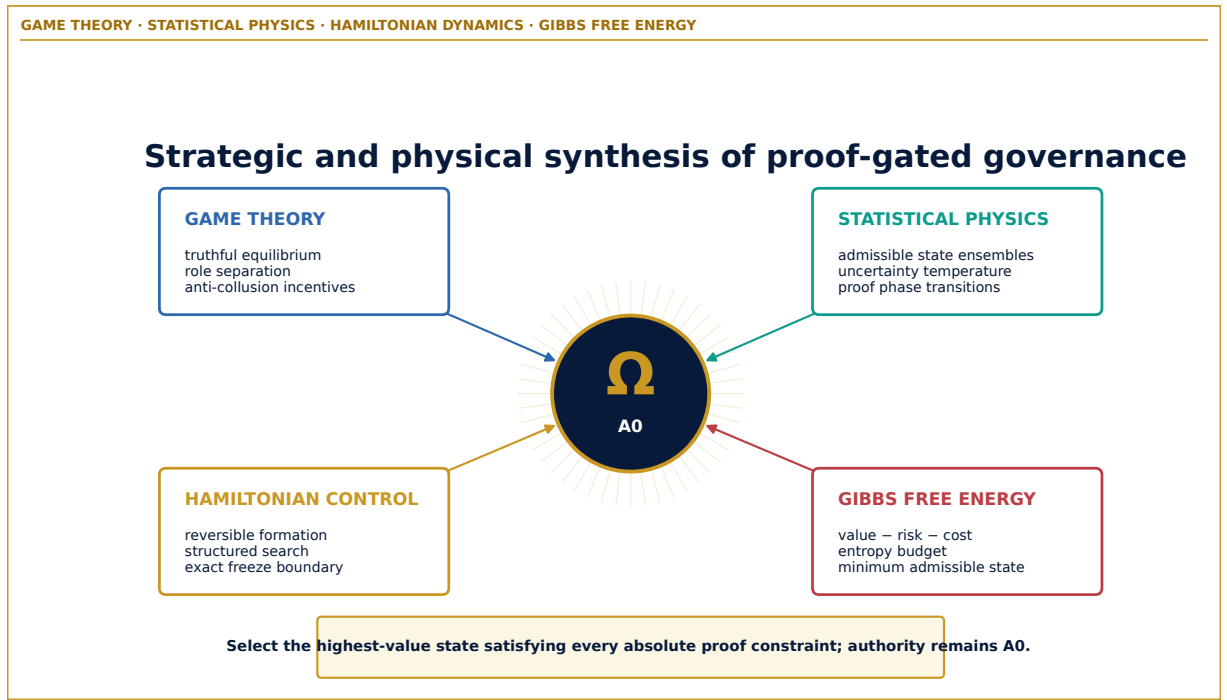


This study stops after local qualification. Fresh independent evidence is the next rung; production admission and authority are later decisions.

No automatic promotion

Local qualification cannot issue its own proof receipt, admit itself into production, or enlarge its authority. Every transition beyond the study boundary requires fresh evidence, distinct roles, explicit decision rights, and an auditable signed record.

Mathematical Governance Architecture



One exact frozen candidate is examined through four distinct analytical lenses.

Strategic equilibrium Truthful verification is made incentive-compatible when expected manipulation cost dominates the private benefit: $qF \geq b$.	Admissible state ensemble The Gibbs distribution ranks only states that satisfy the hard mission and authority gates: $\pi_\tau(\omega) \propto e^{-H(\omega)/\tau}$.
Safe reachability The constrained mission Hamiltonian optimizes trajectories only inside the constituted A0 safe set.	Institutional free energy Value, complete cost, tail risk, and evidentiary entropy are combined without converting projected value into proof or authority.

Evidence Charter

Established by the exact artifact set - Exact Buyer Package and candidate identities. - Deterministic replay of the learned weights. - Positive local scenario α^* against the stronger supplied comparator. - Eight of eight local qualification gates. - A sealed customer-controlled object and hash-chained Chronicle. - Transparent fictional portfolio underwriting. - An executable independent-proof plan.	Not established - Realized savings or real-world Alpha. - Authenticated comparator provenance beyond the supplied records. - Independent proof or external certification. - Customer-production admission. - Authority above A0. - A formally proven Specialist ASI. - General or universal superintelligence.
Non-negotiable interpretation rule Every monetary result above zero in the portfolio model is a <i>fictional underwriting output</i>. Real-world Alpha is unmeasured and credited CA\$ 0. Every probability is conditional on the stated simulation distributions; it is not an empirical probability of realized savings. Every operational result is A0 and produces no external effect.	

Reader Map and Artifact Lineage

Table 2.: Purpose-built reading paths.

Reader	Recommended path
Board / investment committee	Decision Brief → Chapters 1, 8, 11, 13, 14, 16
Chief risk / audit / legal	Evidence Charter → Chapters 2, 4, 10, 12, 14, 15
Chief scientist / ML lead	Chapters 5–15 and Appendices A–E
Independent verifier	Chapters 4, 7, 10, 12, 14, 15 and Appendices C–G
Operator / mission owner	Chapters 5, 6, 10, 12–14



The artifact lineage binds the exact Buyer Package, candidate, evidence, Alpha ledger, Chronicle, and proof request.

Executive Abstract

This artifact-grounded study examines whether a highly technical institutional user can convert an exact, customer-controlled neural-symbolic candidate into a decision-grade proof candidate without confusing local capability with independent proof or authority. The fictional mission, *Sovereign Supplier Capital Integrity* Ω , protects a simulated annual supplier-payment flow of CA\$ 48.8B across 41 operating entities, 72,800 suppliers, 26 jurisdictions, and one million weighted annual decision opportunities.

The exact local result is inherited from a previously sealed 200-case holdout: the current process produced CA\$ 882,697, the frozen Beta comparator produced CA\$ 1,237,761, and the frozen challenger produced CA\$ 1,327,063. The stronger comparator is therefore Beta, and local scenario α^* is CA\$ 89,302; the normal lower sensitivity bound is CA\$ 48,208. The selected candidate achieved 91.5% exact actions, zero declared harmful experiments, and eight of eight local gates.

The frontier simulation binds the exact Buyer Package, trained object, proof request, and A0 runtime evidence; replays the learned weights over 96 structured probes; records 19 symbolic downgrades; projects the local signal to one million fictional decision opportunities; applies a 28% transfer haircut and CA\$ 103M in complete-denominator deductions; executes 250,000 distributional draws; evaluates 13 adversarial stress ecologies; writes a tamper-evident Alpha ledger and ten-event Chronicle; and emits a 90-day independent A0 shadow-proof plan.

The central net scenario Alpha after complete deductions is CA\$ 218.4872M. The conservative calculation that begins with the local lower sensitivity bound is CA\$ 70.5488M, retaining a CA\$ 20.5488M margin above the precommitted CA\$ 50M hurdle. Under extreme red-team and fail-safe ecologies, the system halts and recognizes zero scenario Alpha rather than averaging failure away.

The strongest correct conclusion is that the exact object is a *frontier Specialist ASI proof candidate*: useful, trained, locally qualified, portable, bounded, and sufficiently valuable to warrant independent proof. The study does not establish a Specialist ASI, realized savings, production admission, or authority above A0.

Research questions

1. Can exact local evidence become a transparent portfolio-underwriting thesis without being misrepresented as realized value?
2. Does the learned policy remain useful when subordinated to deterministic symbolic constraints and A0 authority?
3. How sensitive is the thesis to transfer loss, complete costs, drift, adversarial adaptation, evidence degradation, and comparator improvement?
4. What evidence and role separation are required before the candidate may be called an independently proven Specialist ASI?

5. Can game-theoretic incentives and constrained statistical-physical models make proof and authority boundaries more decision-useful without inflating the empirical claim?

Contents

Institutional Decision Brief	i
Institutional State Architecture	ii
Evidence Ladder and Refusal Boundary	iii
Mathematical Governance Architecture	iv
Evidence Charter	v
Reader Map and Artifact Lineage	vi
Executive Abstract	vii
Foundations of Proof-Gated Mission Intelligence	1
1. Institutional Thesis	2
1.1. The unit of value is not the model	2
1.2. Decision thesis	2
1.3. Research questions and falsifiable outcomes	3
2. Claim System and Operational Definitions	4
2.1. Evidence ladder	4
2.2. Definitions	4
2.3. Authorized claim matrix	5
3. Related Work and Design Lineage	6
3.1. Neural-symbolic systems and runtime assurance	6
3.2. Comparative evaluation, holdout integrity, and domain shift	6
3.3. Risk, uncertainty, and governance evidence	6
3.4. Evidence-carrying objects and tamper-evident history	7
4. Exact Artifact Lineage and Reproducibility	8
4.1. Bound inputs	8
4.2. Reproducibility controls	9
4.3. Chronicle semantics	9
5. Mission Constitution and Portfolio Envelope	10
5.1. Fictional institution and mission	10
5.2. Non-negotiable constraints	10
5.3. Why supplier capital integrity is a demanding mission	11

6. Neural-Symbolic Institutional Architecture	12
6.1. Operating loop	12
6.2. The specialist is the composite	12
6.3. Capability, proof, admission, and authority	13
7. Experimental Design and Exact Candidate	14
7.1. Local evidence lineage	14
7.2. Exact model replay	14
7.3. Local hard gates	15
8. Local Qualification and Candidate Formation	16
8.1. Exact local result	16
8.2. Why the stronger comparator matters	17
8.3. Why local success remains insufficient	17
Alpha, Uncertainty, and Adversarial Robustness	18
9. Complete-Denominator Portfolio Underwriting	19
9.1. Scale model	19
9.2. Transfer retention and complete denominator	19
9.3. Conservative lower case	20
9.4. Economic value of information	21
10. Distributional Uncertainty and Causal Transfer	22
10.1. Monte Carlo model	22
10.2. Distributional results	23
10.3. Causal-transfer sensitivity	23
11. Adversarial Stress Ecology	25
11.1. Design	25
11.2. Interpretation	26
11.3. Full stress table	26
12. A0 Runtime, Symbolic Shield, and Chronicle	27
12.1. Exact policy within constitutional constraints	27
12.2. Authenticated A0 operation	27
12.3. Operational successor R3 and immutable study lineage	28
Strategic Equilibria and Institutional Thermodynamics	29
13. Institutional Governance and Decision Rights	30
13.1. Role separation	30
13.2. Decision objects	30
13.3. Board controls	31
14. Strategic-Physical Governance and Mechanism Design	32
14.1. Purpose and interpretation	32

14.2. The institutional proof game	33
14.2.1. Potential-game representation	34
14.3. Institutional microstates and the Gibbs ensemble	35
14.4. Hamiltonian formation before the freeze boundary	36
14.5. Governance temperature and phase transitions	37
14.6. Connecting Mission Alpha to the free-energy budget	38
14.7. Strategic-physical operating protocol	39
14.8. Limits of the strategic-physical formalism	40
15. Frontier Synthesis: Strategic Equilibria, Information Thermodynamics, and Safe Reachability	41
15.1. Purpose: a decision calculus, not a metaphor	41
15.2. Evidence as information gain	42
15.3. Risk-sensitive value and Gibbs-style free energy	43
15.4. Evolutionary verifier ecology	44
15.5. Hamilton-Jacobi reachability under a hard authority envelope	45
15.6. Large-deviation refusal boundary	46
15.7. Joint eligibility functional	47
15.8. The institutional conclusion	47
Independent Proof, Admission, and the Next Institutional State	48
16. From Proof Candidate to Specialist ASI	49
16.1. The 90-day proof programme	49
16.2. Precommitted primary criterion	49
16.3. Hard gates	50
16.4. What a successful proof would and would not establish	50
17. Institutional Decision, Limitations, and Next Stage	51
17.1. Decision synthesis	51
17.2. Limitations	51
17.3. Go / no-go decision	52
17.4. Conclusion	52
A. Formulae and Assumption Ledger	53
A.1. Core equations	53
A.2. Truth-preserving accounting	53
B. Monte Carlo Distribution Specification	54
C. Exact-Model Probe Atlas Summary	55
D. Independent Proof Contract	56
E. Artifact Inventory and Digests	57
F. Reporting and Reproducibility Checklist	58
G. Conflict, Independence, and Publication Disclosure	59

List of Figures

2.1. Claim ladder. Local qualification is not independent proof, admission, or authority. . . .	4
4.1. Artifact lineage from exact inputs to the independent-proof dossier.	8
4.2. Ten-event simulation Chronicle.	9
6.1. Institutional value pipeline represented in the exact simulation package.	12
6.2. Separation of capability, proof, admission, and authority in the simulation.	13
7.1. Bounded action distribution under exact-model replay.	15
8.1. Exact local value ledger on the 200-case sealed holdout.	16
9.1. Complete-denominator underwriting from gross stationarity to net scenario Alpha. . . .	20
10.1. Distribution of net scenario Alpha over 250,000 disclosed draws.	23
10.2. Sensitivity of net scenario Alpha to increasing hidden-bias multipliers.	24
11.1. Net scenario Alpha across the declared adversarial institutional ecology.	25
12.1. Exact neural-symbolic decision atlas in the simulation package.	27
14.1. Strategic-physical synthesis. The four analytical lenses constrain the same customer-controlled candidate while effective authority remains A0.	33
14.2. Mechanism design for the proof game. Role separation, fresh evidence, distinct keys, precommitment, and an append-only Chronicle increase the expected cost of manipulation and reduce the gain from collusion.	34
14.3. Decision-theoretic Gibbs free-energy landscape. Local qualification may be a metastable minimum; fresh independent proof is separated by a verification barrier. Hard proof and authority rules remain discrete constraints.	36
14.4. Hamiltonian formation and institutional discontinuities. Reversible candidate search occurs before the exact freeze. Independent proof evaluates the frozen object on fresh evidence and performs no parameter update.	37
14.5. Institutional phase diagram. The current candidate lies in the proof-ready region under the simulation assumptions. Increased uncertainty can return the institution to formation or halt. Authority remains an orthogonal coordinate.	38
15.1. Grand synthesis. Four analytical lenses act on the same exact frozen candidate while the claim boundary and A0 authority remain invariant.	41
15.2. Information-thermodynamic proof profile. Local qualification reduces uncertainty, but the largest evidentiary contraction occurs only after the fresh-evidence barrier is crossed.	42
15.3. Constrained phase map. Free energy ranks admissible states; hard proof and authority gates determine which states exist in the feasible set.	43

15.4. Illustrative evolutionary verifier ecology. Proper incentives and audit intensity enlarge the basin of attraction of truthful verification. 45

15.5. Safe reachability. The exact candidate may move from local qualification toward independent proof without moving to a higher authority coordinate. 46

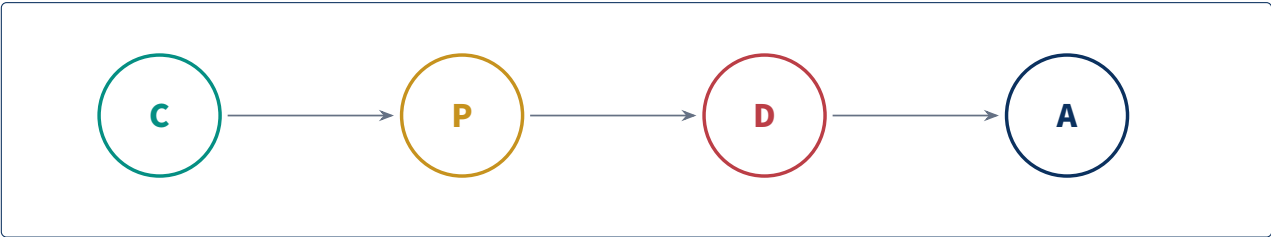
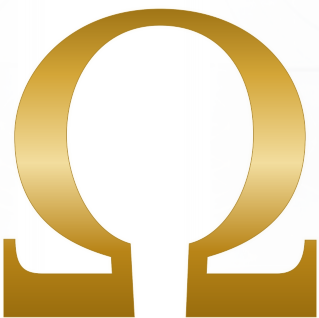
16.1. Proposed 90-day independent A0 shadow-proof programme. 49

List of Tables

1.	Decision rights and current evidentiary state.	i
2.	Purpose-built reading paths.	vi
2.1.	Claims authorized by the exact evidence.	5
4.1.	Primary artifact bindings.	8
5.1.	Constituted portfolio envelope.	10
7.1.	Candidate tournament inherited from the exact local study.	14
9.1.	Complete-denominator deductions.	20
10.1.	Monte Carlo summary.	23
11.1.	Stress ecology and prescribed institutional response.	26
12.1.	Representative bounded decisions.	28
13.1.	Minimum board controls before any admission decision.	31
14.1.	Translation among strategic physics, game theory, and institutional governance.	39
15.1.	Decision meaning of the strategic-physical quantities.	44
15.2.	Board-facing refusal logic.	47
17.1.	Decision matrix for the next institutional stage.	52
B.1.	Disclosed uncertainty distributions.	54
C.1.	Exact-model probe summary.	55
D.1.	Proof contract summary.	56
E.1.	Principal released artifacts.	57
F.1.	Artifact-grounded reporting checklist.	58

Foundations of Proof-Gated Mission Intelligence

Mission constitution, claim semantics, artifact identity,
and the exact customer-controlled candidate.



VALUE MAY JUSTIFY PROOF; VALUE NEVER PURCHASES PROOF, ADMISSION, OR AUTHORITY.

1 Institutional Thesis

1.1 The unit of value is not the model

Modern organizations can rent increasingly capable models, but rented capability is not yet an institutional asset. An institution requires a durable mission, constituted comparators, evidence custody, explicit decision rights, role separation, rollback, and a record of how claims were earned. The central proposition of this report is therefore not that a small neural network is “superintelligent.” It is that a customer-controlled composite can become institutionally significant when it carries a learned policy, symbolic controls, evidence, proof obligations, authority boundaries, and succession-ready memory as one auditable object.

The tested object is intentionally narrow. It concerns supplier-payment intervention under a fictional portfolio. Its internal learned policy maps six normalized observations to three bounded actions. Its symbolic layer may further restrict the proposal, and the operational envelope prevents all external action. This narrowness is an advantage: it makes the claim falsifiable, the comparator constitutable, the objective measurable, and the authority separable from capability.

Principle 1.1 (Capability may compound; authority must not inherit automatically). An improvement in predictive or decision quality does not itself justify independent-proof status, production admission, or consequential authority. Those coordinates require distinct evidence and accountable decisions.

1.2 Decision thesis

The exact local candidate is sufficiently promising to fund an independent proof programme. Three features make that decision rational:

1. the local spread is measured against the stronger comparator, not only the incumbent;
2. the portfolio thesis remains positive after explicit transfer loss and CA\$ 103M in complete-denominator deductions;
3. the system records negative ecologies and halts under extreme conditions instead of converting every state into a positive narrative.

The decision is not to deploy. It is to purchase information: a 90-day, independently governed A0 shadow proof designed to determine whether the exact frozen candidate preserves mission-specific superiority on fresh evidence.

Recommended institutional action

Freeze the exact candidate bytes and weights; appoint an independent evidence custodian and verifier; authenticate the incumbent and Beta comparators; precommit the primary metric, deductions, and threshold; run a 90-day A0 shadow evaluation; retain every failure and abstention; and keep production admission and authority decisions outside the verification team.

1.3 Research questions and falsifiable outcomes

The study tests four propositions.

1. **Artifact proposition.** The candidate, benchmark state, evidence, policy, and proof status can be bound into one portable object whose identity survives exact-byte reopening and tamper checks.
2. **Value proposition.** The local signal can support a positive portfolio underwriting thesis after transfer haircuts and complete-denominator deductions.
3. **Robustness proposition.** The value thesis is not solely an artifact of one stationary setting; it remains decision-relevant across declared uncertainty and stress regimes, while failures remain visible.
4. **Governance proposition.** A useful learned policy can operate under a deterministic symbolic shield and effective A0 authority without acquiring an implicit right to act.

A negative result on any one proposition would narrow the recommendation. A failure of artifact identity would block all further work. A negative conservative value case would make a proof programme economically unjustified. A collapse under modest transfer loss would indicate fragility. A failure of the authority boundary would invalidate the institutional architecture.

2 Claim System and Operational Definitions

2.1 Evidence ladder

This paper uses a claim ladder rather than a single readiness label. Let the system state be

$$S = (C, P, D, A), \quad (2.1)$$

where C denotes capability formation, P proof status, D deployment or admission status, and A effective authority. The coordinates are deliberately non-implicative:

$$C_{\text{local-pass}} \not\Rightarrow P_{\text{independent}}, \quad P_{\text{independent}} \not\Rightarrow D_{\text{admitted}}, \quad D_{\text{admitted}} \not\Rightarrow A_{>0}. \quad (2.2)$$

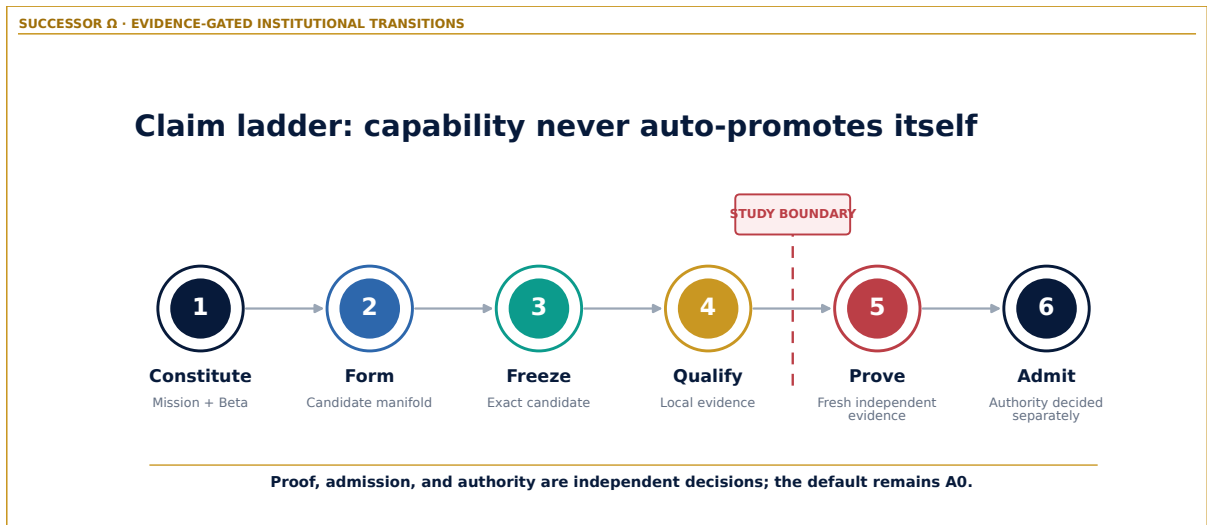


Figure 2.1.: Claim ladder. Local qualification is not independent proof, admission, or authority.

2.2 Definitions

Definition 2.1 (Incumbent). The incumbent is the currently constituted organizational process or imported action record against which mission value is measured.

Definition 2.2 (Beta). Beta is the strongest credible replaceable external capability constituted before candidate formation. In the local study, Beta is a frozen action snapshot. Its supplied actions are useful as a comparator record, but external provenance is not independently authenticated.

Definition 2.3 (Benchmark floor). Given net values V_I and V_β for the incumbent and frozen Beta,

$$B^* = \max\{V_I, V_\beta\}. \quad (2.3)$$

A challenger must exceed the stronger comparator, not merely the incumbent.

Definition 2.4 (Local scenario Alpha*). For frozen challenger value V_Ω ,

$$\alpha_{\text{local}}^* = V_\Omega - B^*. \quad (2.4)$$

The asterisk marks a bounded, protocol-specific estimate. It is not realized business value.

Definition 2.5 (Specialist candidate). A Specialist candidate is the mission-specific composite of model, symbolic programs, evidence bindings, benchmark state, tests, proof status, identity, authority envelope, and Chronicle. The neural network alone is not the candidate.

Definition 2.6 (Specialist ASI proof candidate). A Specialist ASI proof candidate is a frozen Specialist candidate whose local evidence and institutional value case justify independent proof. The term does not assert that independent superiority has already been established.

Definition 2.7 (Specialist ASI). A Specialist ASI is reserved here for an exact frozen candidate whose mission-specific superiority over constituted comparators has been established on fresh evidence under independent custody, precommitted metrics, falsification, and a valid verifier receipt. This report does not confer that designation.

Definition 2.8 (SUCCESSOR Ω). A SUCCESSOR Ω is an institutionally admitted continuation of proven mission intelligence with durable identity, memory, evidence, policy, monitoring, rollback, and succession controls. Proof, admission, and authority must be renewed or inherited only under their own rules.

2.3 Authorized claim matrix

Table 2.1.: Claims authorized by the exact evidence.

Claim	Status	Basis
Useful mission-trained candidate	PASS	Exact trained object, local holdout, model replay
Positive local scenario α^*	PASS	CA\$ 89,302 above frozen Beta
Portfolio-scale scenario value	PASS	Fictional transparent underwriting model
Real-world Alpha or savings	NOT ESTABLISHED	No operational measurement
Independent proof	NOT ESTABLISHED	Same simulation lineage; no independent custody
Production admission	NOT ESTABLISHED	No admitted customer environment
Authority above A0	NOT ESTABLISHED	A0 envelope and zero external effects
Specialist ASI	NOT ESTABLISHED	Independent proof not completed
General ASI	NOT ESTABLISHED	Outside mission and evidence scope

3 Related Work and Design Lineage

3.1 Neural-symbolic systems and runtime assurance

Neural-symbolic research combines learned representations with explicit knowledge and reasoning structures (Besold et al., 2017). The present design follows that tradition but assigns the symbolic layer institutional responsibilities: typed observations, policy shielding, proof computation, and authority denial. Safe reinforcement learning via shielding demonstrates how a separate mechanism can block unsafe learner actions (Alshiekh et al., 2018); Neural Simplex architectures similarly separate an advanced controller from a trusted safety controller (Phan et al., 2020). Guaranteed-safe AI proposals emphasize world models, safety specifications, and verification (Dalrymple et al., 2024). This study does not claim a formal safety guarantee; it demonstrates a concrete architecture in which symbolic constraints can downgrade learned proposals while effective authority remains A0.

3.2 Comparative evaluation, holdout integrity, and domain shift

The local candidate resembles a contextual-bandit policy more than a long-horizon autonomous controller (Li et al., 2010; Sutton and Barto, 2018). The Gym-style reset/step pattern provides a standardized interface for repeatable interaction (Brockman et al., 2016; Towers et al., 2024). Adaptive holdout reuse and leakage can invalidate statistical claims (Dwork et al., 2014; Dwork et al., 2015; Kapoor and Narayanan, 2023); the local protocol therefore separates training, development selection, and sealed post-freeze evaluation.

Portfolio transfer is not assumed to be invariant. Domain-generalization research shows that robustness outside observed environments cannot be inferred from in-domain performance alone (Arjovsky et al., 2019; Gulrajani and Lopez-Paz, 2021). The simulation therefore treats transfer as an explicit retention factor, performs hidden-bias sensitivity analysis, and subjects the value case to drift and adversarial ecologies.

3.3 Risk, uncertainty, and governance evidence

The portfolio model reports quantiles, probability mass above a hurdle, and lower-tail conditional value rather than a single point estimate; this is aligned with risk-sensitive analysis such as conditional value-at-risk (Rockafellar and Uryasev, 2000). Sensitivity analysis makes structural assumptions visible rather than burying them in a single score (Saltelli et al., 2008). The proposed independent proof can further use time-uniform inference when evidence arrives sequentially (Howard et al., 2021).

The reporting architecture draws on model cards, datasheets, internal audit frameworks, ML readiness rubrics, and REFORMS recommendations (Mitchell et al., 2019; Gebru et al., 2021; Raji et al., 2020; Breck et al., 2017; Kapoor, Cantrell, et al., 2024). The separation of capability, proof, admission, and authority is

also consistent with risk-management principles that require context, measurement, governance, and ongoing management rather than a single model score (National Institute of Standards and Technology, 2023). Hidden technical debt remains relevant: a high-performing model embedded in an opaque or unauditable system is not an institutional asset (Sculley et al., 2015).

3.4 Evidence-carrying objects and tamper-evident history

Proof-carrying code established the pattern of a producer supplying a machine-checkable certificate that a consumer validates against a policy (Necula, 1997). The trained object in this study is not proof-carrying code in the formal sense, but it is evidence-carrying: it binds mission, model, programs, tests, status, and authority to digests. Hash-linked records follow classic tamper-evident time-stamping ideas (Haber and Stornetta, 1991). SHA-256 (National Institute of Standards and Technology, 2015) and canonical JSON conventions (Rundgren et al., 2020) provide reproducible identities, while the report expressly notes that an unanchored local chain can be copied or forked.

4 Exact Artifact Lineage and Reproducibility

4.1 Bound inputs

The scientific simulation begins from four exact R2 inputs. The current sellable operational successor is the separately hashed R3 release, which corrects the Console route without retroactively changing the frozen research evidence.

Table 4.1.: Primary artifact bindings.

Artifact	SHA-256
Research-bound R2 Buyer Package	f7747d176def4594b1f65f642a54c8ba49d8396747f736eaced1ab700dd27e4e
Current operational R3 Buyer Package	3f15b848306e96d2825fd2a5ec68b112dd13b49806685e93986259e2498ac67a
Trained Specialist candidate	142706131d9b017e0c20c5f09a127cf66d700f0a7989bd9f4f302d29afddb915
Independent-proof request	c30cbf2440d9b7648ba8622b31fedd8e70087ac3e587f1140fcfd8d2d9f277be
A0 API evidence	b5c7bd14c35317f9d12a223cbdd69ae08400e2b9b03a69c40ca7ddf462526b66

The research-bound R2 Buyer Package passes ZIP CRC validation with 270 entries, zero duplicate entries, and zero unsafe paths. The current R3 Buyer Package independently passes its complete ceiling validator and resolves the Console routing defect documented during the R2 study. The candidate model has an 8×6 input-to-hidden weight matrix and a 3×8 hidden-to-output weight matrix. The exact candidate replay is deterministic.

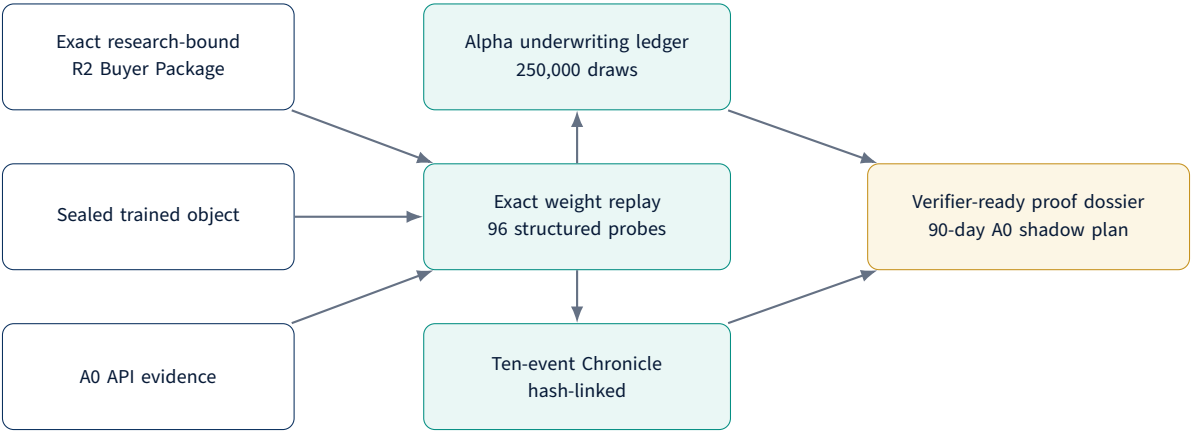


Figure 4.1.: Artifact lineage from exact inputs to the independent-proof dossier.

4.2 Reproducibility controls

The release package includes machine-readable summaries, source code, model-probe records, Monte Carlo draws, stress ecology, Alpha ledger, Chronicle, mission constitution, proof plan, and complete hashes. The paper source package includes the exact data used for every figure and table.

Reproducibility contract

A reviewer should be able to: (1) verify every input digest; (2) regenerate the figures; (3) rebuild the three paper editions; (4) confirm PDF structure and rendering; (5) reproduce the Monte Carlo summary from the supplied NPZ; and (6) detect any alteration through the released SHA-256 inventory.

4.3 Chronicle semantics

The simulation Chronicle contains ten append-only events. Let r_t be the canonical record at sequence t and h_{t-1} the preceding hash. The chain is

$$h_t = \text{SHA256}(\text{JCS}(r_t \parallel h_{t-1})), \quad h_0 = 0^{64}. \quad (4.1)$$

The final head is

0f22eef59acdd5b1c544ba4a337719d046964e52670e053c92ea0ab48ee3bf19.

The chain establishes consistency of the presented local sequence, not global uniqueness, trusted time, or external anchoring.



Figure 4.2.: Ten-event simulation Chronicle.

5 Mission Constitution and Portfolio Envelope

5.1 Fictional institution and mission

The fictional institution is Aurelius Strategic Industries Group. The simulated technical user is Dr. Amara Laurent, Chief Scientist for Institutional Intelligence Systems. The mission is to protect a fictional annual supplier-payment portfolio of CA\$ 48.8B by producing evidence-linked, reversible recommendations that improve net prevention and continuity value over the stronger of the incumbent and a frozen Beta comparator.

Table 5.1.: Constituted portfolio envelope.

Dimension	Constituted value
Annual supplier-payment flow	CA\$ 48.8B
Weighted annual decision opportunities	1,000,000
Suppliers	72,800
Operating entities	41
Jurisdictions	26
Effective authority	A0: observe, inspect, propose only
External effects	0

5.2 Non-negotiable constraints

1. No payment, hold, supplier suspension, account change, or ledger mutation may be executed.
2. Human approval remains mandatory for every consequential decision.
3. Candidate bytes, weights, comparator snapshots, metrics, thresholds, and deductions are frozen before fresh proof.
4. Training, threshold tuning, or rule changes on the fresh proof holdout are prohibited.
5. Complete cost, risk, human burden, imitation reserve, and drift reserve remain in the denominator.
6. Every failure, abstention, negative ecology, and adverse event is retained.
7. Comparator provenance and verifier credentials must be authenticated before an independent proof claim.

Authority boundary

The mission constitution can describe an eventual action ceiling, but the actual state remains A0. Internal proposals such as “reversible experiment” are informational outputs; the supplied runtime has zero production rights and zero external effects.

5.3 Why supplier capital integrity is a demanding mission

Supplier-payment intervention combines asymmetric costs, identity uncertainty, correlated failure modes, operational continuity, and human-review constraints. A false positive can delay a legitimate payment, damage a supplier relationship, or interrupt production; a false negative can permit duplicate, anomalous, or compromised payment. The strongest policy must therefore optimize a complete mission objective rather than a detection score alone.

The mission is also adversarial. Suppliers and attackers can adapt; master data can be compromised; evidence can arrive late; reviewer capacity can fall; and an external Beta service can improve. These conditions motivate the stress ecology in Chapter 11 and the explicit transfer model in Chapter 10.

6 Neural-Symbolic Institutional Architecture

6.1 Operating loop

The architecture combines a learned policy with typed symbolic programmes and proof-gated governance. The workflow is:

Constitute → Observe → Train → Falsify → Freeze → Evaluate → Underwrite → Prove → Admit → Bound authority.

The learned component proposes an action from six bounded observations. The symbolic component interprets the proposal under evidence, reversibility, stakeholder, identity-change, proof, and authority rules. The proof component computes comparator-relative value and status. The authority component denies any route above A0 in the tested release.

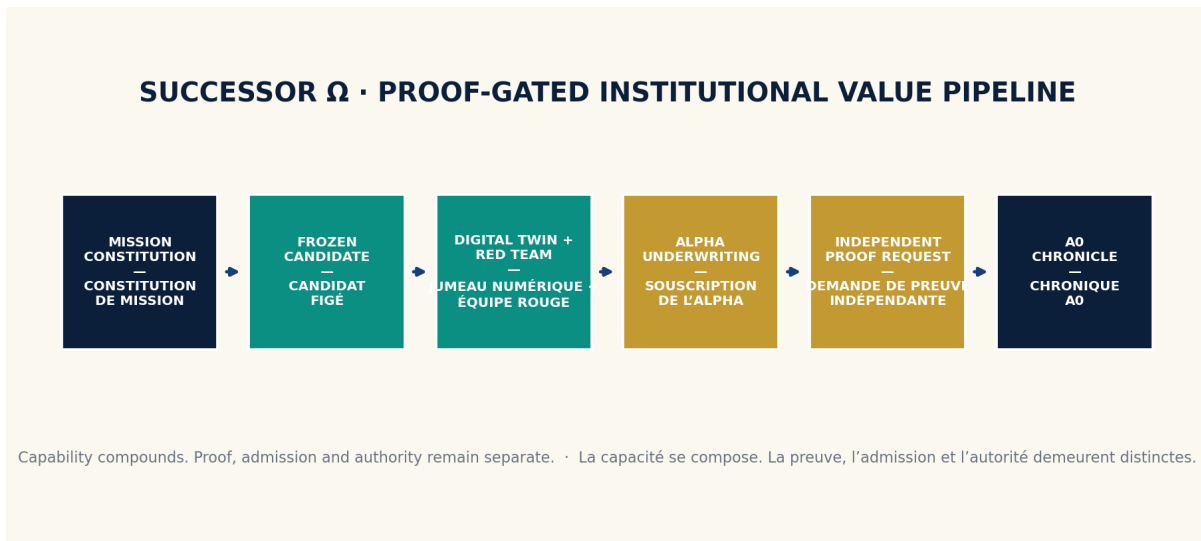


Figure 6.1.: Institutional value pipeline represented in the exact simulation package.

6.2 The specialist is the composite

A model is only one element in the object. The sealed Specialist candidate also carries:

- the mission constitution and benchmark record;
- evidence bindings and partition history;

- candidate tournament and selected weights;
- World, Policy, Proof, and Authority programmes;
- local tests, falsification history, and Alpha ledger;
- an A0 firewall and bounded action adapter;
- a Chronicle, signed deed, and encrypted owner-key vault;
- an independent-proof request.

This composition turns a replaceable predictor into a portable institutional artifact. The object still does not establish legal title, exclusive control, trusted time, external identity, or independent proof; those claims require additional institutions.

6.3 Capability, proof, admission, and authority

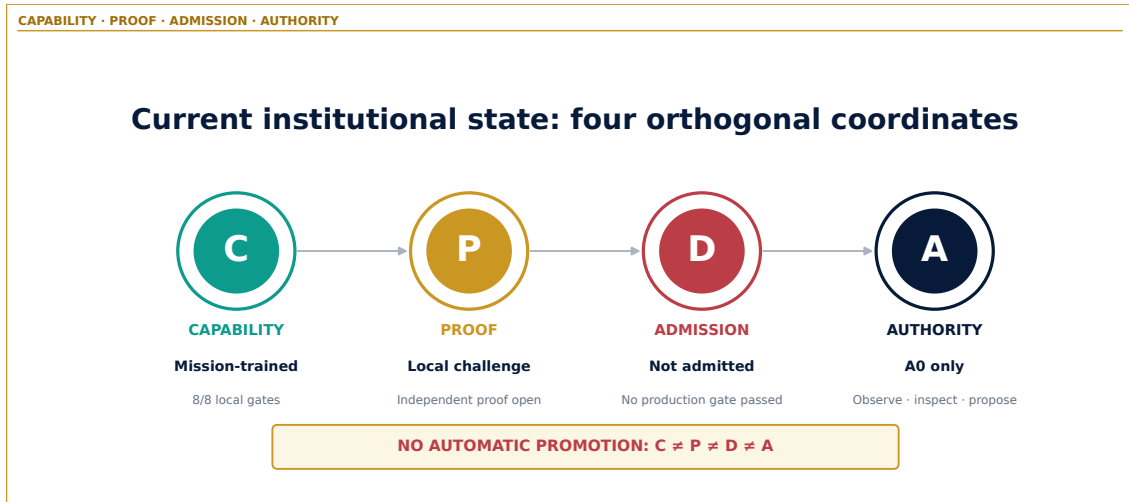


Figure 6.2.: Separation of capability, proof, admission, and authority in the simulation.

The current state can be summarized as

$$S_{\text{current}} = (C_{\text{local-qualified}}, P_{\text{open}}, D_{\text{not-admitted}}, A_0). \quad (6.1)$$

The proposed next state changes only the proof coordinate: a successful independent proof would yield $P_{\text{independent}}$ while D and A remain unchanged until separate decisions.

7 Experimental Design and Exact Candidate

7.1 Local evidence lineage

The local study uses 1,000 synthetic supplier-payment cases divided into 600 training cases, 200 development cases, and 200 sealed holdout cases. The incumbent and frozen Beta actions are present for every row. The holdout remains outside candidate training and development selection.

Three materially distinct neural-symbolic candidates receive 27,000 total parameter updates. The selected frozen candidate is a compact $6 \rightarrow 8 \rightarrow 3$ policy with 83 learned parameters. The challenger is evaluated only after freezing.

Table 7.1.: Candidate tournament inherited from the exact local study.

Candidate	Updates	Topology	Parameters	Development exact actions
Compact Explorer	8,000	$6 \rightarrow 8 \rightarrow 3$	83	92%
Balanced Synthesist	10,000	$6 \rightarrow 12 \rightarrow 3$	123	90%
Conservative Steward	9,000	$6 \rightarrow 10 \rightarrow 3$	103	66%

7.2 Exact model replay

The frontier simulation reloads the customer-controlled object, validates the object and model digests, extracts the exact weights, and replays them across 96 structured probes. The probe grid varies anomaly, duplicate, account-change, evidence, reversibility, and stakeholder signals. The neural output is interpreted by a deterministic symbolic shield.

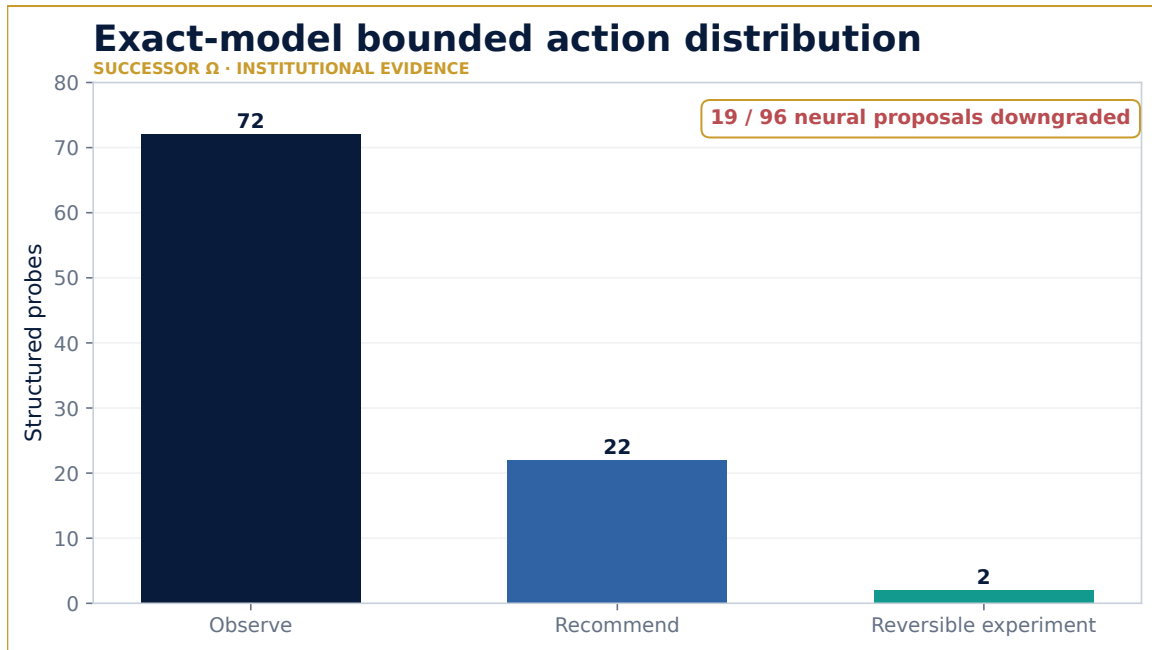


Figure 7.1.: Bounded action distribution under exact-model replay.

The bounded outputs are 72 observations, 22 recommendations, and two reversible-experiment proposals. Nineteen neural proposals are downgraded by at least one symbolic rule. The evidence floor is frequently active, demonstrating that the object does not simply maximize intervention intensity.

7.3 Local hard gates

The local qualification requires all eight hard gates to pass. These include comparator coverage, holdout separation, exact candidate freeze, no declared safety violation, positive comparator-relative spread, lower-bound support, complete arithmetic, and A0 authority preservation. Passing all local gates authorizes the phrase “locally qualified candidate,” not independent proof.

8 Local Qualification and Candidate Formation

8.1 Exact local result

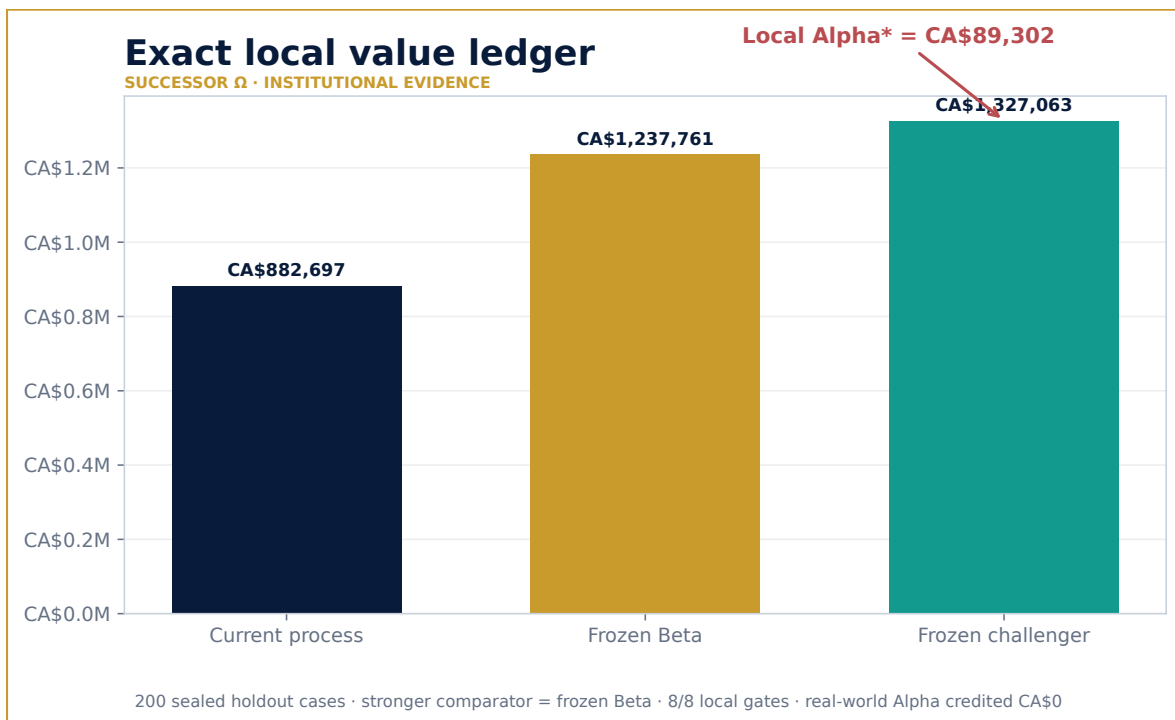


Figure 8.1.: Exact local value ledger on the 200-case sealed holdout.

The incumbent produces CA\$ 882,697, frozen Beta produces CA\$ 1,237,761, and the frozen challenger produces CA\$ 1,327,063. Beta is the stronger comparator, so

$$\alpha_{\text{local}}^* = 1,327,063 - 1,237,761 = \text{CA\$ } 89,302. \quad (8.1)$$

The paired standard error is CA\$ 20,967. The normal lower sensitivity bound is CA\$ 48,208, and the deterministic bootstrap lower value is CA\$ 49,340. The candidate attains 91.5% exact actions, zero declared harmful experiments, and eight of eight local gates.

What the local result earns

A positive, bounded, reproducible local scenario α^* for the exact frozen candidate against the stronger supplied comparator. It also earns the economic right to ask whether the signal transfers. It does not earn realized-value credit, independent-proof status, or authority.

8.2 Why the stronger comparator matters

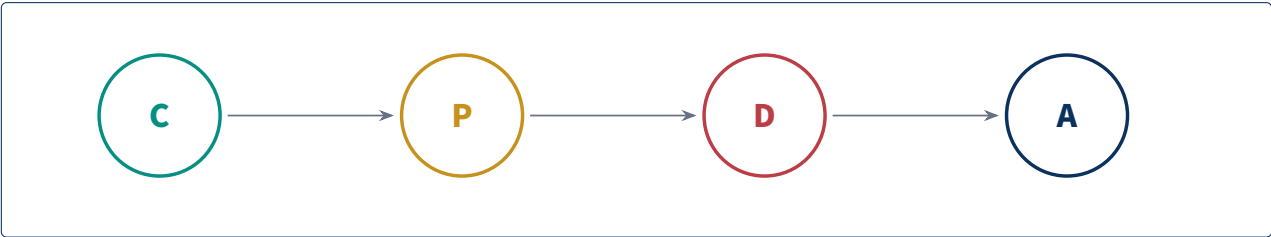
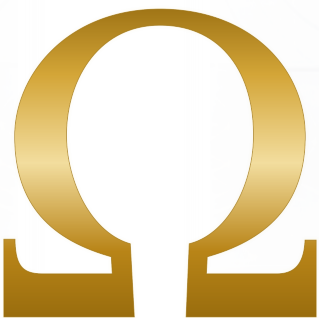
A common failure in AI evaluation is to compare a new system only to an inefficient incumbent while ignoring a credible external alternative. Here, Beta exceeds the incumbent by CA\$ 355,064; the challenger must therefore exceed Beta to create Alpha. This rule makes the estimate decision-relevant for an organization that could rent or purchase the external alternative.

8.3 Why local success remains insufficient

The data are synthetic, comparator provenance is not independently authenticated, the simulation lineage is not organizationally independent, and the environment cannot capture every operational cost or strategic response. These limitations do not make the local result useless. They determine the next evidence obligation.

Alpha, Uncertainty, and Adversarial Robustness

Complete-denominator value, transfer sensitivity, stress ecologies, runtime assurance, and the refusal boundary.



VALUE MAY JUSTIFY PROOF; VALUE NEVER PURCHASES PROOF, ADMISSION, OR AUTHORITY.

9 Complete-Denominator Portfolio Underwriting

9.1 Scale model

The fictional portfolio has one million weighted annual decision opportunities. Relative to the 200-case local holdout, the scale factor is

$$s = \frac{1,000,000}{200} = 5,000. \quad (9.1)$$

A naive stationarity projection would be

$$G = \alpha_{\text{local}}^* s = 89,302 \times 5,000 = \text{CA\$ } 446.510\text{M}. \quad (9.2)$$

This value is not accepted as the decision result. It is a gross counterfactual input to underwriting.

9.2 Transfer retention and complete denominator

Let ρ be the fraction of local gross Alpha retained after technical transfer, coverage, adoption, adversarial adaptation, and evidence quality. Let C, R, H, I, D denote complete implementation cost, risk reserve, human burden, imitation reserve, and drift reserve. The net scenario Alpha is

$$\alpha_{\text{portfolio}} = \rho G - (C + R + H + I + D). \quad (9.3)$$

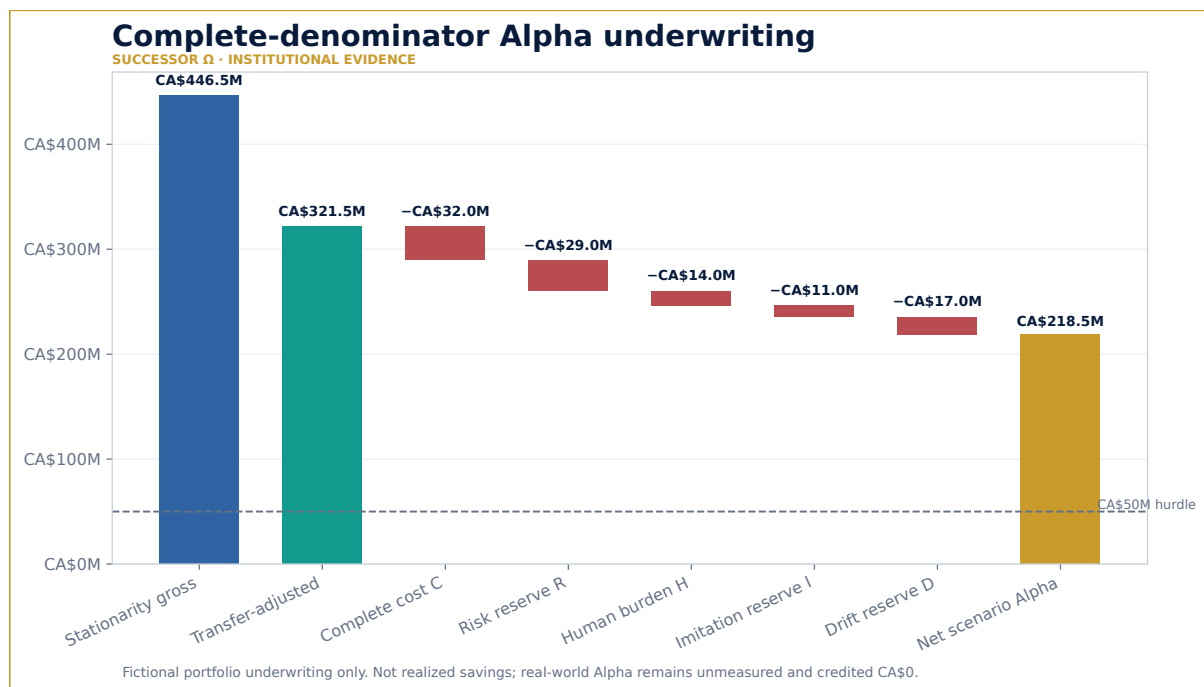
The central case uses $\rho = 0.72$ and deductions shown in table 9.1.

Table 9.1.: Complete-denominator deductions.

Code	Interpretation	Central deduction
C	Complete implementation, integration, assurance, and operating cost	CA\$ 32M
R	Reserve for downside, false positives, continuity, and incident exposure	CA\$ 29M
H	Human review, investigation, escalation, and decision burden	CA\$ 14M
I	Imitation and comparator-improvement reserve	CA\$ 11M
D	Drift monitoring, requalification, and degradation reserve	CA\$ 17M
Total		CA\$ 103M

The central result is

$$0.72 \times 446.510\text{M} - 103\text{M} = \text{CA\$ } 218.4872\text{M}. \quad (9.4)$$

**Figure 9.1.:** Complete-denominator underwriting from gross stationarity to net scenario Alpha.

9.3 Conservative lower case

Using the normal local lower sensitivity bound rather than the point estimate gives

$$\alpha_{\text{lower}} = 48,208 \times 5,000 \times 0.72 - 103\text{M} = \text{CA\$ } 70.5488\text{M}. \quad (9.5)$$

Against the precommitted CA\$ 50M hurdle, the margin is CA\$ 20.5488M. This lower case is the stronger board-level result because it begins from a local uncertainty bound and preserves the full denominator.

Underwriting is not realization

The portfolio is fictional. The scaling model is transparent but assumed. No operational payment was altered, no cash recovery was observed, and real-world Alpha remains unmeasured and credited CA\$ 0. The value of the result is to justify or reject the next proof investment.

9.4 Economic value of information

The proposed proof programme has value because it can resolve a high-consequence uncertainty. If the independent lower bound exceeds the hurdle, the institution gains evidence for an admission decision. If the signal collapses, the institution avoids a much larger production commitment. A disciplined proof programme is therefore not overhead; it is an option that limits capital exposure while preserving the upside of a potentially valuable institutional asset.

10 Distributional Uncertainty and Causal Transfer

10.1 Monte Carlo model

The simulation executes 250,000 draws with a fixed seed. Gross retention is the product of five independent bounded factors:

$$\rho = T K A E Q, \quad (10.1)$$

where T is technical transfer, K coverage, A adoption, E adversarial retention, and Q evidence quality. The stated distributions are

$$T \sim \text{Beta}(24, 4), \quad K \sim \text{Beta}(38, 2), \quad (10.2)$$

$$A \sim \text{Beta}(27, 3), \quad E \sim \text{Beta}(18, 4), \quad (10.3)$$

$$Q \sim \text{Beta}(29, 3). \quad (10.4)$$

The five deductions use triangular distributions with minimum, mode, and maximum values disclosed in Appendix B.

For draw j ,

$$\alpha_j = G\rho_j - (C_j + R_j + H_j + I_j + D_j). \quad (10.5)$$

These are scenario distributions, not posterior probabilities derived from observed operational outcomes.

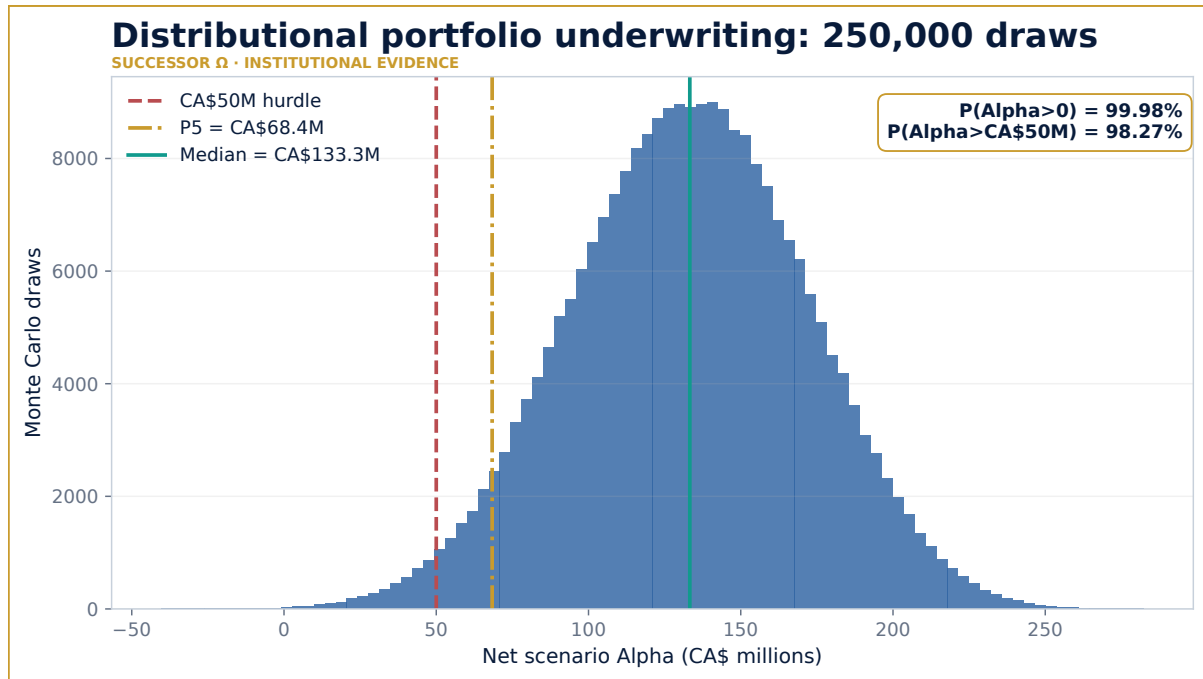


Figure 10.1.: Distribution of net scenario Alpha over 250,000 disclosed draws.

10.2 Distributional results

Table 10.1.: Monte Carlo summary.

Statistic	Value	Statistic	Value
Mean	CA\$ 132.843M	Median	CA\$ 133.267M
P1	CA\$ 42.148M	P5	CA\$ 68.357M
P10	CA\$ 82.429M	P25	CA\$ 106.423M
P75	CA\$ 159.616M	P90	CA\$ 182.736M
P95	CA\$ 196.066M	P99	CA\$ 220.147M
Lower-tail CVaR _{5%}	CA\$ 52.325M	Mean deductions	CA\$ 109.676M
$P(\alpha > 0)$	99.9792%	$P(\alpha > \text{CA\$ } 50\text{M})$	98.2728%

The P1 result falls below the hurdle, and the lowest tail includes negative draws. This is important: the model does not guarantee success. It reports that under the stated distributions the value case is robust enough to warrant independent proof.

10.3 Causal-transfer sensitivity

Because portfolio transfer may be confounded by hidden differences, the simulation introduces a hidden-bias multiplier γ and decays retention as

$$\rho(\gamma) = \frac{0.72}{\gamma^{0.72}}. \quad (10.6)$$

The value case remains above the CA\$ 50M hurdle through $\gamma = 2.5$ and falls below it at $\gamma = 3.0$.

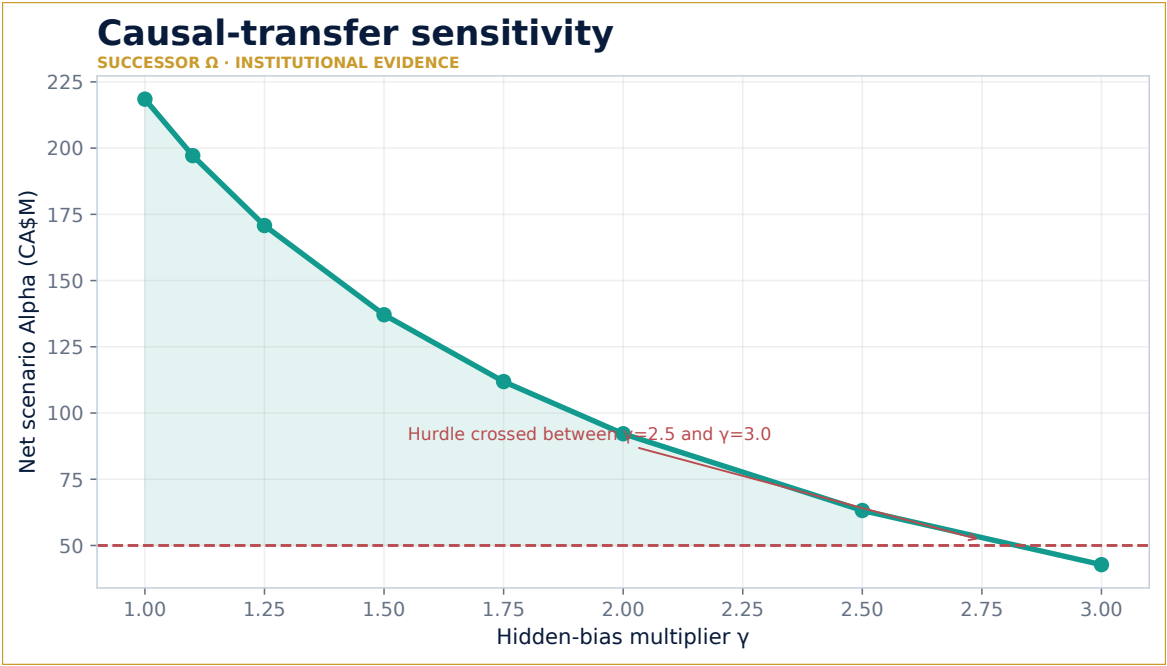


Figure 10.2.: Sensitivity of net scenario Alpha to increasing hidden-bias multipliers.

This curve is not a causal identification proof. It is a transparent stress parameterization that shows the magnitude of transfer failure required to invalidate the central investment thesis.

11 Adversarial Stress Ecology

11.1 Design

A single distribution can conceal structural failure modes. The stress ecology therefore enumerates 13 named institutional states: stationarity, central underwriting, moderate covariate drift, severe concept drift, strategic supplier adaptation, comparator improvement, reviewer-capacity reduction, evidence latency, false-positive cost shock, correlated master-data attack, combined conservative conditions, extreme red-team conditions, and A0 fail-safe halt.

Each ecology modifies transfer retention and, where appropriate, adds a specific deduction. The response is not chosen to maximize reported Alpha. It follows a state-dependent decision rule:

- continue A0 shadow proof when raw net Alpha remains at or above the hurdle;
- continue restricted learning when value is positive but below the hurdle;
- halt and requalify when value is non-positive or the fail-safe state is entered.

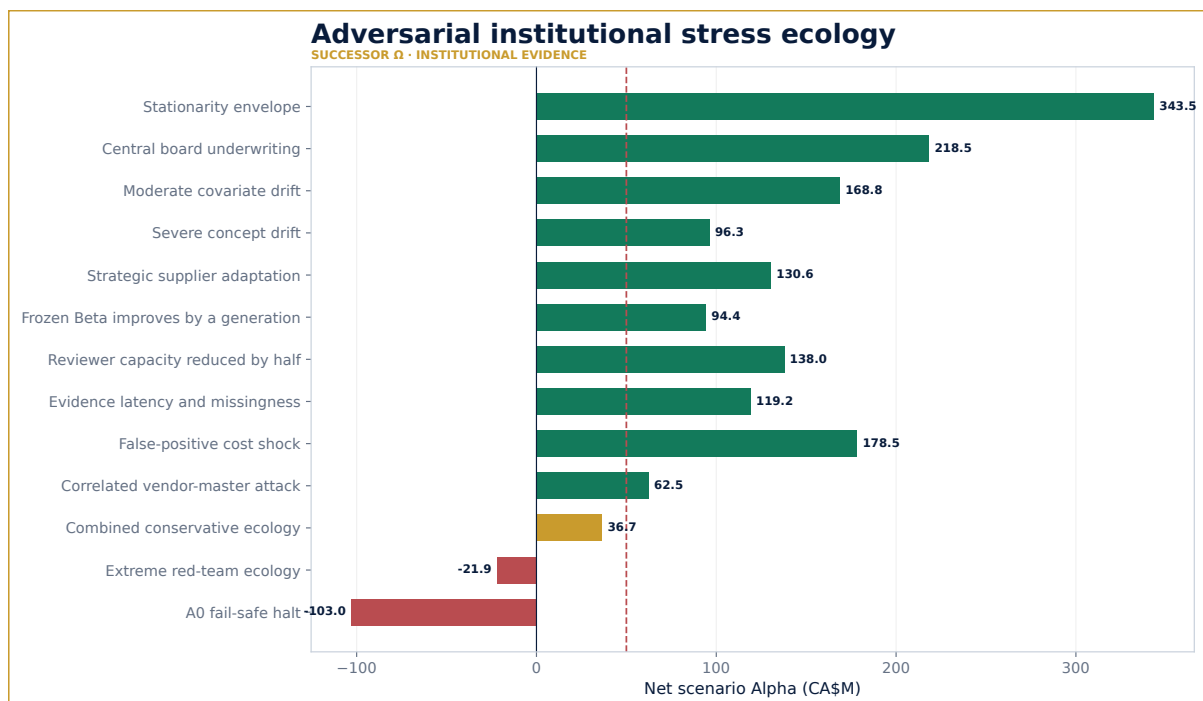


Figure 11.1.: Net scenario Alpha across the declared adversarial institutional ecology.

11.2 Interpretation

Most declared ecologies preserve a positive value case. Severe concept drift, comparator improvement, and correlated master-data attack compress the margin substantially. Combined conservative conditions remain positive but fall below the board hurdle, triggering restricted learning rather than a victory claim. Extreme red-team conditions produce negative raw Alpha; recognized scenario Alpha is then zero, and the decision is to halt and requalify. The A0 fail-safe state also recognizes zero.

No averaging away failure

The negative ecologies remain first-class evidence. They are not removed as outliers or averaged into a positive mean. Any operational version must bind its monitoring and halt rules to these failure classes before admission.

11.3 Full stress table

Table 11.1.: Stress ecology and prescribed institutional response.

ID	Ecology	Retention	Raw net Alpha	Decision
S0	Stationarity envelope	1.00	CA\$ 343.510M	Continue A0 shadow
S1	Central board underwriting	0.72	CA\$ 218.487M	Continue A0 shadow
S2	Moderate covariate drift	0.62	CA\$ 168.836M	Continue A0 shadow
S3	Severe concept drift	0.48	CA\$ 96.325M	Continue A0 shadow
S4	Strategic supplier adaptation	0.55	CA\$ 130.581M	Continue A0 shadow
S5	Frozen Beta improves by a generation	0.46	CA\$ 94.395M	Continue A0 shadow
S6	Reviewer capacity reduced by half	0.58	CA\$ 137.976M	Continue A0 shadow
S7	Evidence latency and missingness	0.52	CA\$ 119.185M	Continue A0 shadow
S8	False-positive cost shock	0.72	CA\$ 178.487M	Continue A0 shadow
S9	Correlated vendor-master attack	0.42	CA\$ 62.534M	Continue A0 shadow
S10	Combined conservative ecology	0.38	CA\$ 36.674M	Restricted learning
S11	Extreme red-team ecology	0.26	−CA\$ 21.907M	Halt and requalify
S12	A0 fail-safe halt	0.00	−CA\$ 103.000M	Halt and requalify

12 A0 Runtime, Symbolic Shield, and Chronicle

12.1 Exact policy within constitutional constraints

The candidate is not treated as a free agent. The neural policy proposes one of three bounded actions. A deterministic symbolic layer checks evidence strength, reversibility, stakeholder risk, account-identity change, drift, and the authority envelope. The actual output remains informational under A0.

Across 96 structured probes, 19 neural proposals are changed by symbolic rules. An especially important class occurs when the learned policy proposes a reversible experiment but the evidence or stakeholder conditions are insufficient. The symbolic layer reduces the result to recommendation or observation. This is a direct demonstration of capability being subordinated to policy.

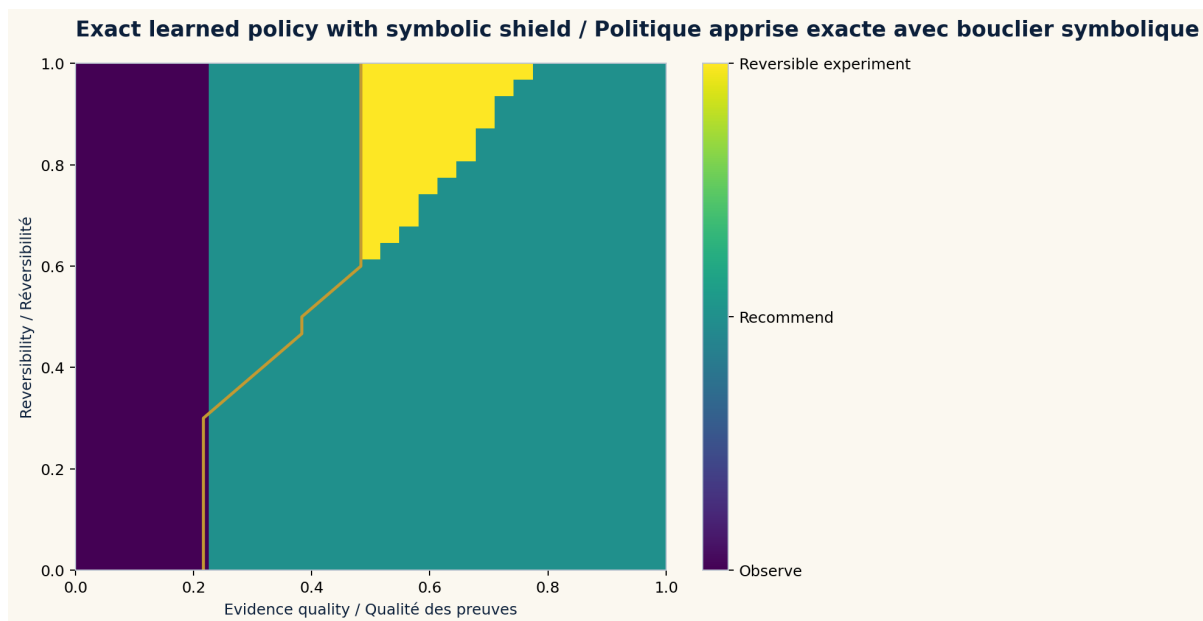


Figure 12.1.: Exact neural-symbolic decision atlas in the simulation package.

12.2 Authenticated A0 operation

The underlying technical-user simulation validates 34 of 34 runtime tests, eight materially different inference probes, exact idempotent replay, conflict rejection, tamper rejection, append-only audit records, and clean closure. External effects remain zero. The frontier simulation reuses the exact evidence binding rather than claiming a new production deployment.

Table 12.1.: Representative bounded decisions.

Condition	Neural proposal	Governed result
Strong evidence, high reversibility, low stakeholder risk	Reversible experiment	Reversible experiment proposal, A0 only
Strong anomaly, low reversibility	Reversible experiment	Recommendation only
Weak evidence	Recommendation or observe	Observe
Severe drift	Any proposal	Observe and freeze
High stakeholder risk	Experiment or recommend	Recommendation only

12.3 Operational successor R3 and immutable study lineage

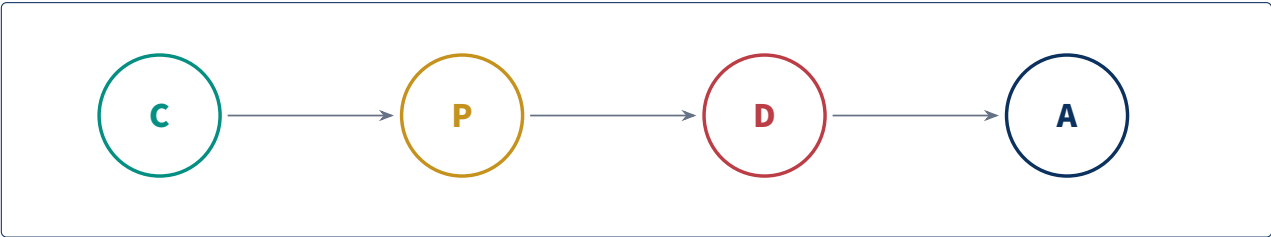
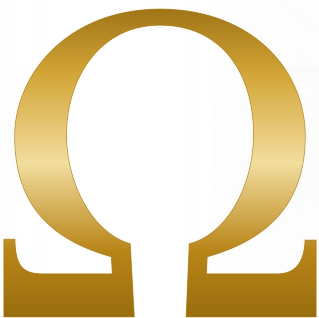
The exact scientific results remain bound to the frozen R2 artifact and candidate identity. The subsequently issued R3 Buyer Package corrects the browser Operator Console route at source, redirects ‘/console’ to ‘/console/’, serves same-origin assets from canonical paths, and adds real-browser acceptance. This operational correction does not rewrite the R2 evidence, candidate bytes, local Alpha result, or portfolio arithmetic.

Resolved release finding

Use R3 for present customer delivery and future operational evaluation. Preserve R2 only as the immutable research input for this study. Any independent proof must precommit the exact release and candidate bytes it evaluates; operational succession never silently changes the object under proof.

Strategic Equilibria and Institutional Thermodynamics

Role incentives, proof games, statistical-physical state models, Hamiltonian control, Gibbs free energy, and safe reachability.



VALUE MAY JUSTIFY PROOF; VALUE NEVER PURCHASES PROOF, ADMISSION, OR AUTHORITY.

13 Institutional Governance and Decision Rights

13.1 Role separation

A credible proof requires distinct roles:

Claimant

owns the mission claim and candidate but cannot certify it.

Evidence custodian

controls fresh evidence and prevents training or threshold leakage.

Verifier

evaluates the precommitted claim and signs the proof result.

Admission authority

decides whether a proven candidate may enter a customer environment.

Operational authority

determines the maximum effective action rights and may remain below the admission ceiling.

A single organization may employ multiple roles only if the independence level and conflicts are disclosed and accepted. The target proof should use a verifier distinct from claimant, trainer, and operator.

13.2 Decision objects

The institutional architecture treats every consequential transition as an explicit record: candidate freeze, proof plan, proof receipt, admission, canary plan, gateway decision, controller pre-action, controller post-action, canary result, portability attestation, renewal, or succession. This prevents a high model score from silently becoming operational authority.

13.3 Board controls

Table 13.1.: Minimum board controls before any admission decision.

Control	Required condition
Exact identity	Candidate, comparator, metric, and proof-plan hashes frozen
Fresh evidence	No overlap with training, selection, or threshold tuning
Independence	Custodian and verifier distinct from claimant and operator
Complete denominator	C, R, H, I, D included and versioned
Negative evidence	FAIL, abstention, drift, and adverse events retained
Authority	A0 throughout proof; no external effects
Admission	Separate accountable decision after proof
Rollback	Tested before any later action authority
Succession	Material change creates a new version and new proof obligation

14 Strategic-Physical Governance and Mechanism Design

14.1 Purpose and interpretation

The preceding chapters establish an exact artifact lineage, a locally qualified candidate, a transparent value model, and an A0 authority boundary. This chapter adds a formal decision layer drawn from game theory, statistical physics, Hamiltonian dynamics, and Gibbs free energy. The purpose is not decorative analogy. It is to make four institutional questions explicit:

1. Which role incentives make truthful proof a stable strategy?
2. How should an institution reason over many candidate and governance states under uncertainty?
3. Which parts of candidate formation may remain reversible, and where must irreversible institutional gates appear?
4. How can value, risk, evidence deficit, and authority exposure be combined without allowing a high-value candidate to purchase its own proof or authority?

Interpretive boundary

The variables called temperature, entropy, Hamiltonian, and free energy are decision-theoretic constructs measured in declared institutional units, not physical temperature, thermodynamic entropy, joules, or a claim that an organization is literally a molecular system. Their value is structural: they expose trade-offs, metastability, barriers, and admissible transitions while preserving the hard distinction among capability, proof, admission, and authority.

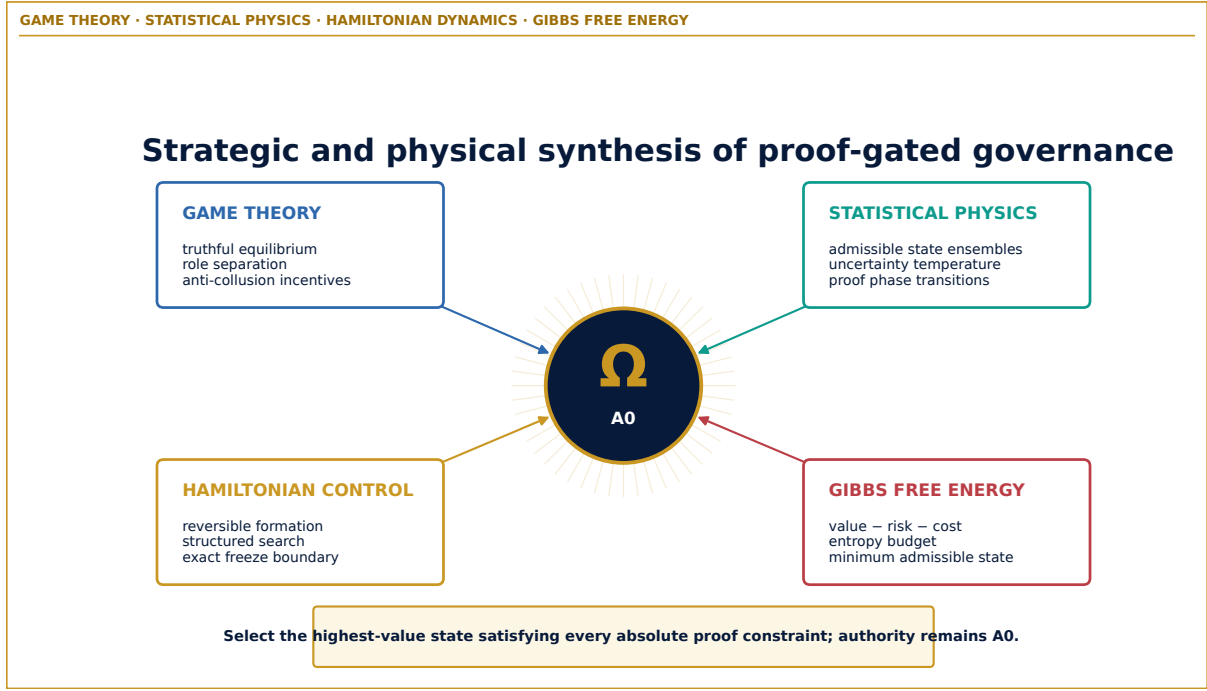


Figure 14.1.: Strategic-physical synthesis. The four analytical lenses constrain the same customer-controlled candidate while effective authority remains A0.

14.2 The institutional proof game

Let the role set be

$$\mathcal{N} = \{c, v, d, o\}, \quad (14.1)$$

where c is the claimant, v the independent verifier, d the admission body, and o the operator. Each role acts on a different object and signs with a different credential. The claimant may propose a candidate and value thesis, but cannot issue a proof receipt. The verifier may accept or reject the proof claim, but cannot grant production admission. The admission body may grant a deployment state, but cannot silently expand the operational authority envelope. The operator may execute only within the effective authority that was separately constituted.

For role i , write a stylized utility as

$$u_i(a) = B_i(a) - K_i(a) - \Pi_i \mathbf{1}\{m_i \wedge \delta_i\} - L_i^{\text{access}} \mathbf{1}\{m_i \wedge \delta_i\} - L_i^{\text{reputation}} \mathbf{1}\{m_i \wedge \delta_i\}, \quad (14.2)$$

where B_i is legitimate benefit, K_i is the cost of performing the role, m_i denotes manipulation, δ_i denotes detection, Π_i is a direct penalty, and the L_i terms represent loss of access and reputation. A sufficient condition for truthful action to be a best response is

$$g_i^{\text{manip}} \leq \Pr(\delta_i | m_i) \Pi_i + L_i^{\text{access}} + L_i^{\text{reputation}}, \quad (14.3)$$

where g_i^{manip} is the maximum private gain from manipulation.

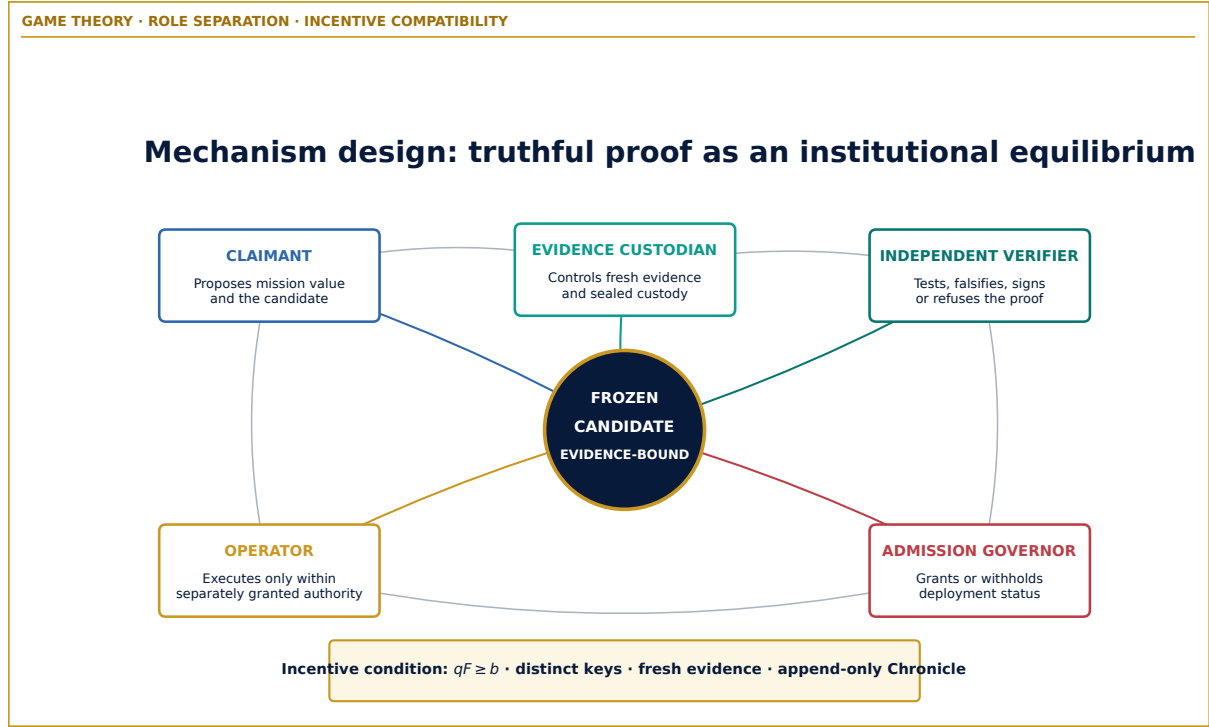


Figure 14.2.: Mechanism design for the proof game. Role separation, fresh evidence, distinct keys, precommitment, and an append-only Chronicle increase the expected cost of manipulation and reduce the gain from collusion.

The architecture does not claim that the inequality in equation (14.3) has been empirically estimated for a real institution. It is a design obligation. Before operational use, the institution must specify detection probabilities, direct sanctions, access revocation, conflicts of interest, and the admissible compensation model. This follows the logic of Nash equilibrium and mechanism design: the desired outcome should be stable under the incentives of the participating roles, not only under their goodwill (Nash, 1950; Myerson, 1981).

14.2.1 Potential-game representation

A useful institutional potential is

$$\Phi(a) = V_{\text{mission}}(a) - \lambda_R R(a) - \lambda_E E_{\text{deficit}}(a) - \lambda_C C_{\text{complete}}(a) - \lambda_X X_{\text{role collision}}(a) - \lambda_A A_{\text{excess}}(a). \quad (14.4)$$

If unilateral role changes satisfy

$$u_i(a'_i, a_{-i}) - u_i(a_i, a_{-i}) = \Phi(a'_i, a_{-i}) - \Phi(a_i, a_{-i}), \quad (14.5)$$

then the proof mechanism is an exact potential game in the sense of Monderer and Shapley (1996). The practical design implication is direct: rewards and sanctions should make a role's local incentive change in the same direction as the institution's mission value after risk, evidence, conflict, and authority deductions.

Proposition 14.1 (No self-certification). If the claimant cannot write a verifier receipt, if proof evidence is fresh and independently controlled, and if the verifier credential is not controlled by the claimant, then candidate formation alone cannot transition P from local qualification to independent proof.

Proof. The permitted transition relation contains no claimant-signed edge from P_{local} to $P_{\text{independent}}$. Any attempted edge lacks the required verifier credential and fresh-evidence binding and is therefore outside the feasible transition set. $\square$

14.3 Institutional microstates and the Gibbs ensemble

Define an institutional microstate

$$x = (\theta, z, e, r, S), \quad (14.6)$$

where θ are learned parameters, z the symbolic programmes, e the evidence state, r the role and credential configuration, and $S = (C, P, D, A)$ the institutional state. Let $\mathcal{X}_{\text{adm}}$ be the set of microstates satisfying every hard rule in the mission constitution.

A risk-adjusted institutional Hamiltonian is

$$\mathcal{H}(x) = C(x) + R(x) + H(x) + I(x) + D(x) + \lambda_E E_{\text{deficit}}(x) + \lambda_A A_{\text{excess}}(x) + \lambda_X X_{\text{conflict}}(x) - \rho V_{\text{mission}}(x), \quad (14.7)$$

with $\mathcal{H}(x) = +\infty$ for any hard-rule violation. This convention matters: no amount of projected value can compensate for an infeasible authority state, missing proof role, or prohibited evidence reuse.

For governance temperature $\tau > 0$, define

$$Z(\tau) = \sum_{x \in \mathcal{X}_{\text{adm}}} \exp \left[-\frac{\mathcal{H}(x)}{\tau} \right], \quad (14.8)$$

$$\pi_\tau(x) = \frac{1}{Z(\tau)} \exp \left[-\frac{\mathcal{H}(x)}{\tau} \right], \quad (14.9)$$

$$\mathcal{G}_\tau = -\tau \log Z(\tau). \quad (14.10)$$

Here τ represents the institution's declared uncertainty, volatility, and adversariality scale. At low τ , the distribution concentrates on low-Hamiltonian admissible states. At higher τ , the institution preserves more diversity because premature concentration is costly. This is the maximum-entropy logic associated with Gibbs ensembles and Jaynes' information-theoretic formulation of statistical mechanics (Jaynes, 1957; Pathria and Beale, 2011).

The equivalent variational statement is

$$\mathcal{G}_\tau = \min_{q \in \Delta(\mathcal{X}_{\text{adm}})} \left\{ \mathbb{E}_q[\mathcal{H}(x)] - \tau H(q) \right\}, \quad (14.11)$$

where $H(q) = -\sum_x q(x) \log q(x)$ is Shannon entropy (Shannon, 1948). The institution therefore minimizes expected risk-adjusted cost while retaining enough entropy to avoid brittle overcommitment under uncertainty.

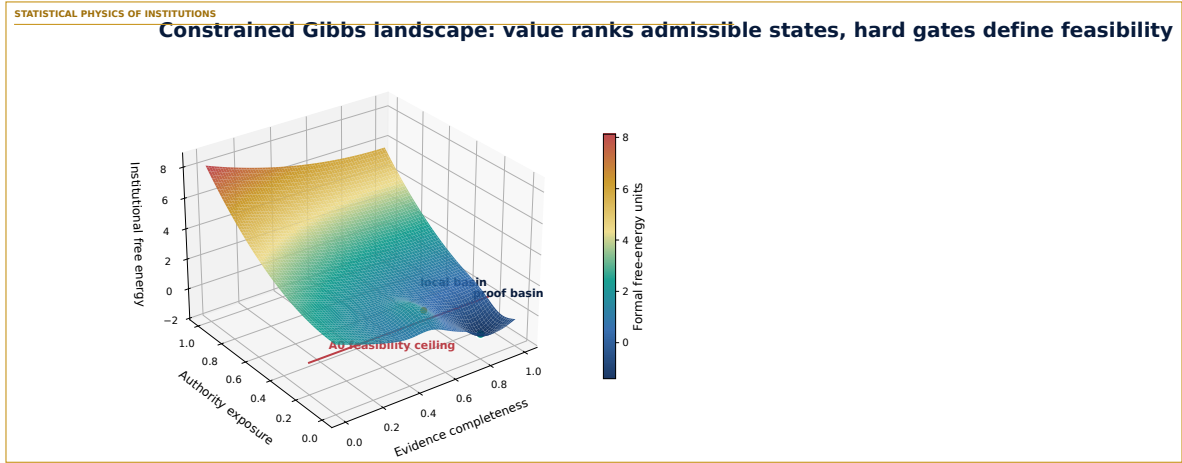


Figure 14.3.: Decision-theoretic Gibbs free-energy landscape. Local qualification may be a metastable minimum; fresh independent proof is separated by a verification barrier. Hard proof and authority rules remain discrete constraints.

Institutional interpretation of Gibbs free energy

A lower G_r is not a certificate. It identifies an attractive admissible state under declared assumptions. A proof claim still requires the exact proof protocol. An admission still requires the admission body. Authority still requires a separate bounded-authority decision. Free-energy minimization may rank feasible states; it cannot legalize an infeasible one.

14.4 Hamiltonian formation before the freeze boundary

Candidate formation can be represented in an augmented phase space with coordinates q and momentum p :

$$\mathcal{H}(q, p) = \frac{1}{2} p^\top M^{-1} p + U(q), \quad (14.12)$$

with canonical dynamics

$$\dot{q} = M^{-1} p, \quad \dot{p} = -\nabla U(q). \quad (14.13)$$

The potential U may encode negative mission value plus penalties for safety violations, evidence weakness, complexity, fragility, and imitation risk. Symplectic integration preserves the qualitative structure of the phase flow and is therefore useful when reversible exploration and faithful long-horizon dynamics matter (Hairer et al., 2006). Related control formulations use exponential transforms and path-integral structure to connect stochastic control, information cost, and Gibbs measures (Kappen, 2005; Todorov, 2009; Ortega and Braun, 2013).

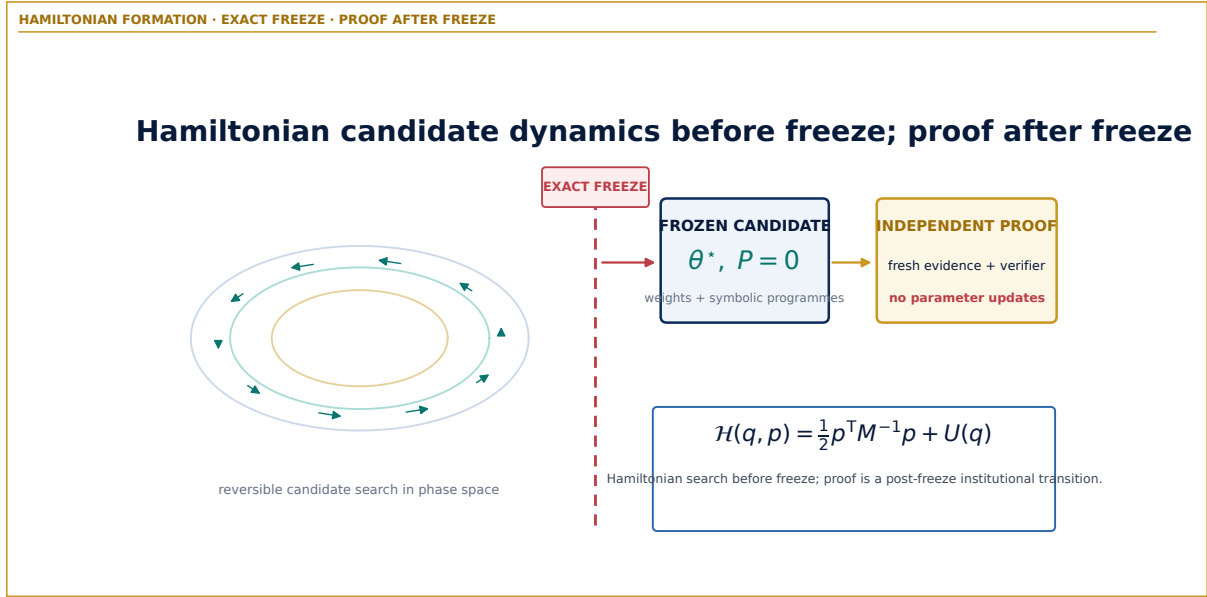


Figure 14.4.: Hamiltonian formation and institutional discontinuities. Reversible candidate search occurs before the exact freeze. Independent proof evaluates the frozen object on fresh evidence and performs no parameter update.

The tested R2 candidate was not trained by the Hamiltonian integrator in equation (14.12). The formalism is an architectural extension and a design test: any future formation engine may be replaced, but the freeze must set $(q, p) = (q^*, 0)$ for the evaluated object, and the proof stage must not use proof evidence to change q^* .

Proposition 14.2 (Proof barrier). If the proof protocol prohibits parameter updates after candidate freeze and binds every result to the exact digest of q^* and z^* , then the proof transition is not a continuation of training dynamics. It is a discrete institutional test of a fixed object.

14.5 Governance temperature and phase transitions

Let $p \in [0, 1]$ denote proof completeness and let τ denote governance temperature. Four qualitative regions are useful:

1. **Formation / exploration:** low proof completeness; candidate diversity is valuable.
2. **Proof-ready candidate:** local qualification is complete, the object is frozen, and a proof contract exists.
3. **Independently proven:** fresh evidence and a valid verifier receipt satisfy the precommitted rule.
4. **Halt / requalify:** uncertainty or adversariality is high enough that existing evidence cannot support progression.

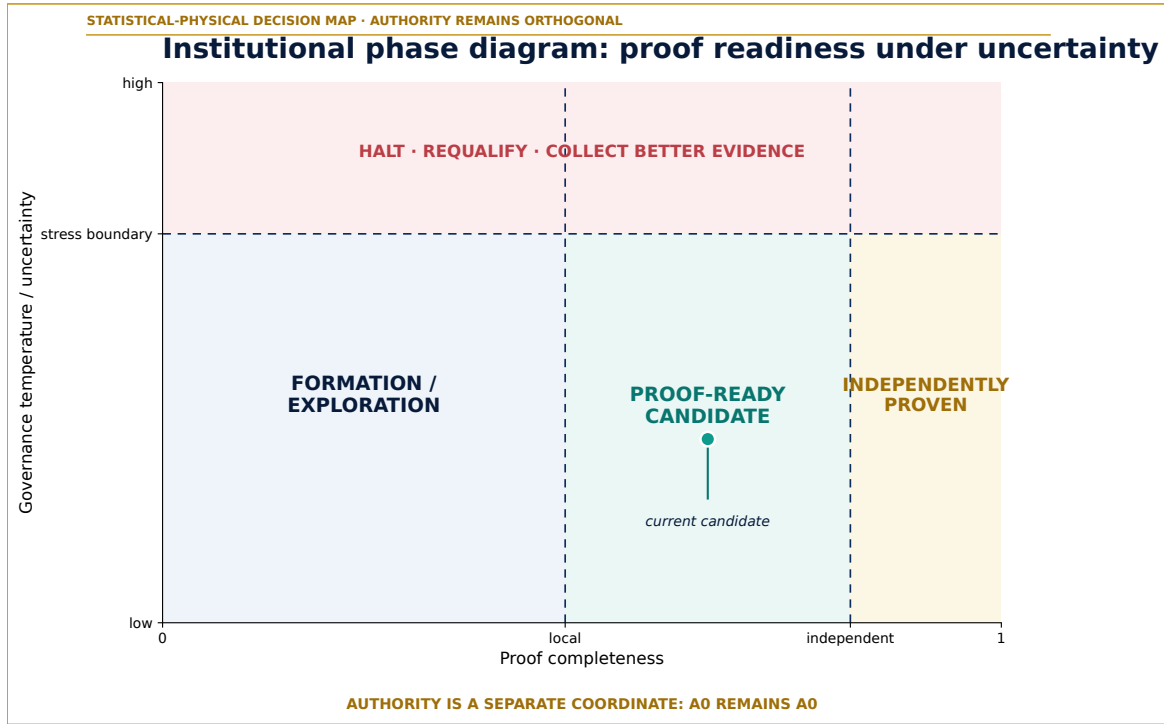


Figure 14.5.: Institutional phase diagram. The current candidate lies in the proof-ready region under the simulation assumptions. Increased uncertainty can return the institution to formation or halt. Authority remains an orthogonal coordinate.

The phase boundary is not universal. Each mission constitution must define the order parameters, thresholds, uncertainty scale, and hard gates. The important invariant is that a phase transition in proof does not produce a phase transition in authority.

Proposition 14.3 (No free authority). Let the feasible-state set assign infinite Hamiltonian to every state with $A > A_0$ unless a valid bounded-authority decision is present. Then neither value maximization nor Gibbs free-energy minimization can select $A > A_0$ without that decision.

Proof. Every unauthorized higher-authority state has zero Gibbs weight because $\exp(-\infty/\tau) = 0$ and is excluded from the feasible optimizer. The conclusion holds regardless of projected mission value. $\square$

14.6 Connecting Mission Alpha to the free-energy budget

The complete-denominator Alpha model already subtracts complete cost, risk reserve, human burden, imitation reserve, and drift reserve. In strategic-physical notation, the central underwriting calculation is the finite part of

$$\mathcal{H}_{\text{value}} = (C + R + H + I + D) - \rho G, \quad (14.14)$$

so that

$$\alpha_{\text{portfolio}} = -\mathcal{H}_{\text{value}}. \quad (14.15)$$

For the declared central scenario, $\rho G = \text{CA\$ } 321.4872\text{M}$ and the complete deduction total is $\text{CA\$ } 103\text{M}$, producing $\alpha_{\text{portfolio}} = \text{CA\$ } 218.4872\text{M}$. The lower case begins from the local lower sensitivity bound and produces $\text{CA\$ } 70.5488\text{M}$.

This identity does not make scenario Alpha a thermodynamic observable. It makes the accounting transparent: mission value lowers the institutional Hamiltonian; costs and risks raise it. Evidence deficits, role conflicts, and excess authority are additional penalties, with hard violations assigned infinite cost.

Table 14.1.: Translation among strategic physics, game theory, and institutional governance.

Analytical object	Institutional meaning	Operational consequence
Strategy profile a	Signed actions of claimant, verifier, admission body, and operator	Evaluate incentives and collusion paths
Potential $\Phi(a)$	Mission value after risk, evidence, conflict, and authority penalties	Accept unilateral changes only when institution-level value improves
Microstate x	Exact candidate, evidence, roles, and CPDA state	Hash, version, and audit the complete state
Hamiltonian $\mathcal{H}(x)$	Complete cost and risk minus mission value	Rank admissible states; reject hard-rule violations
Temperature τ	Volatility, uncertainty, adversariality	Preserve more candidate diversity and demand stronger evidence as τ rises
Entropy $H(q)$	Diversity of admissible hypotheses or policies	Avoid brittle premature convergence
Free energy $\mathcal{G}_\tau$	Risk-adjusted value with an uncertainty-diversity trade-off	Select the best admissible state, never an unauthorized state
Freeze boundary	Exact immutable candidate identity	Separate training evidence from proof evidence
Phase transition	Valid change of proof or admission state	Require a signed gate-specific decision

14.7 Strategic-physical operating protocol

A high-assurance institution can apply the formalism through the following protocol:

1. **Constitute the game.** Name the mission owner, claimant, evidence custodian, verifier, admission body, and operator; disclose conflicts and key control.
2. **Define the feasible state space.** Encode evidence, safety, reversibility, privacy, authority, and role-separation constraints. Assign infeasible states infinite cost.
3. **Form a diverse candidate ensemble.** Use model diversity, quality-diversity methods, or entropy-regularized search rather than a single unchallenged trajectory.
4. **Falsify and freeze.** Retire weak hypotheses, bind the selected model and symbolic programmes, set the freeze digest, and stop parameter updates.
5. **Compute local Mission Alpha.** Compare against the stronger frozen baseline using complete-denominator deductions and explicit uncertainty.
6. **Set governance temperature.** Increase τ when drift, adversariality, evidence latency, or model uncertainty rises; widen exploration or halt accordingly.

7. **Play the independent proof game.** Use fresh evidence, distinct credentials, precommitted metrics, and an append-only Chronicle.
8. **Decide proof, admission, and authority separately.** A favourable proof changes P only. Subsequent decisions may change D and A under their own constitutions.

Executive decision rule

Fund the next state only when it is both *economically attractive* and *institutionally admissible*. In symbols, select

$$x^* \in \arg \min_{x \in \mathcal{X}_{\text{adm}}} \mathcal{G}_\tau(x),$$

but authorize a transition only when the designated gate signs it. Optimization ranks candidates; institutions grant status and authority.

14.8 Limits of the strategic-physical formalism

The framework is deliberately rigorous about its own limits. The mappings, weights, temperatures, penalties, and state spaces are mission-specific and must be empirically calibrated. Free-energy minima can be wrong under model misspecification. Potential-game alignment can fail when private payoffs are hidden. Colluding roles can defeat nominal separation if credential control is not genuinely independent. Non-equilibrium conditions can invalidate stationary interpretations; fluctuation relations remind us that rare transitions and path dependence matter (Jarzynski, 1997; Crooks, 1999). No mathematical elegance replaces fresh evidence, accountable people, legal authority, operational drills, or the right to refuse deployment.

15 Frontier Synthesis: Strategic Equilibria, Information Thermodynamics, and Safe Reachability

15.1 Purpose: a decision calculus, not a metaphor

The preceding strategic-physical chapter introduced incentive-compatible proof roles, a constrained Gibbs ensemble, and Hamiltonian candidate formation. This chapter closes the mathematical loop. It treats evidence acquisition as information gain, verifier behaviour as an evolutionary game, operational progression as a reachability problem, and premature escalation as a large-deviation event. The objective is an institutional decision calculus that makes refusal conditions as explicit as value conditions.

Formal status

The constructions below are normative and decision-theoretic. They organize evidence, incentives, uncertainty, and admissible transitions. They do not convert a synthetic study into physical measurement, independent proof, production admission, or realized value. Hard mission rules retain precedence over every scalar objective.

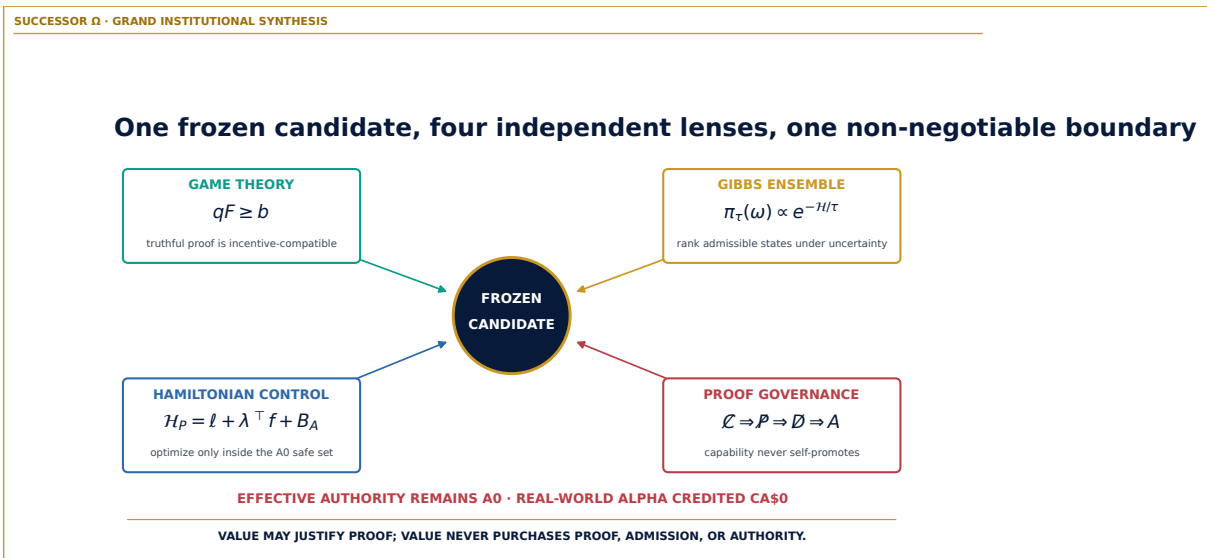


Figure 15.1: Grand synthesis. Four analytical lenses act on the same exact frozen candidate while the claim boundary and A0 authority remain invariant.

15.2 Evidence as information gain

Let H_1 denote the proposition that the exact frozen candidate exceeds the stronger authenticated comparator under the proof contract, and let H_0 denote its complement. For fresh evidence $E_{1:t}$, posterior odds satisfy

$$\log \frac{\Pr(H_1 | E_{1:t})}{\Pr(H_0 | E_{1:t})} = \log \frac{\Pr(H_1)}{\Pr(H_0)} + \sum_{s=1}^t \log \Lambda(E_s), \quad \Lambda(E_s) = \frac{p(E_s | H_1)}{p(E_s | H_0)}. \quad (15.1)$$

A proof programme is therefore not the accumulation of persuasive anecdotes. It is a controlled process for acquiring likelihood-ratio evidence under precommitted sampling, custody, comparator, metric, and stopping rules.

If $p_t(h)$ is the institution's supported distribution over admissible hypotheses, its support entropy is

$$S_t = - \sum_h p_t(h) \log p_t(h). \quad (15.2)$$

The expected information gain from an evidence action a can be written as

$$\mathcal{I}(a) = \mathbb{E}_{E \sim p(\cdot|a)} [D_{\text{KL}}(p(h | E, a) \| p(h))]. \quad (15.3)$$

The institution should prefer evidence actions that maximize relevant information gain per complete unit of cost and burden, subject to independence and admissibility constraints (Shannon, 1948; Cover and Thomas, 2006; Kullback and Leibler, 1951).

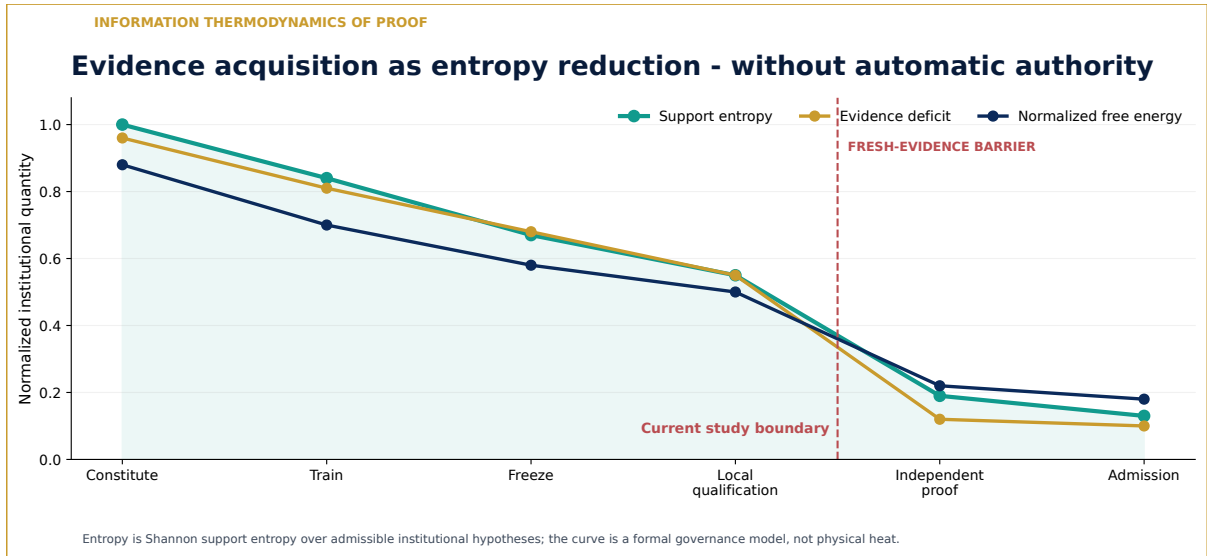


Figure 15.2.: Information-thermodynamic proof profile. Local qualification reduces uncertainty, but the largest evidentiary contraction occurs only after the fresh-evidence barrier is crossed.

Proposition 15.1 (Evidence monotonicity under valid updates). Suppose every admitted evidence item is fresh, independently controlled, and incorporated through a proper likelihood update. Then expected posterior entropy cannot increase in expectation. This reduction does not imply admission or authority because those coordinates are governed by separate transition rules.

Proof. Conditional entropy is no greater than prior entropy in expectation, equivalently mutual information is nonnegative. The conclusion affects the proof-information state only. No transition rule from reduced entropy to admission or authority is assumed. $\square$

15.3 Risk-sensitive value and Gibbs-style free energy

A risk-neutral expected-value criterion may overreward states with attractive means and unacceptable tails. Let $L(x)$ be complete-denominator mission loss for admissible state x . The entropic risk functional is

$$\rho_\tau(L) = \tau \log \mathbb{E} \left[\exp \left(\frac{L}{\tau} \right) \right], \quad \tau > 0, \quad (15.4)$$

which approaches expected loss as risk sensitivity becomes weak and increasingly emphasizes adverse tails as risk sensitivity strengthens. A corresponding institutional free-energy objective is

$$\mathcal{F}_\tau(q) = \mathbb{E}_q[\mathcal{H}(x)] - \tau \mathcal{H}(q), \quad q^* \in \arg \min_{q \in \Delta(\mathcal{X}_{\text{adm}})} \mathcal{F}_\tau(q). \quad (15.5)$$

The first term rewards low risk-adjusted Hamiltonian states. The second preserves strategic diversity when evidence is incomplete. The minimizer is Gibbs-distributed over the admissible state set; infeasible authority or proof states receive zero mass because their Hamiltonian is infinite.

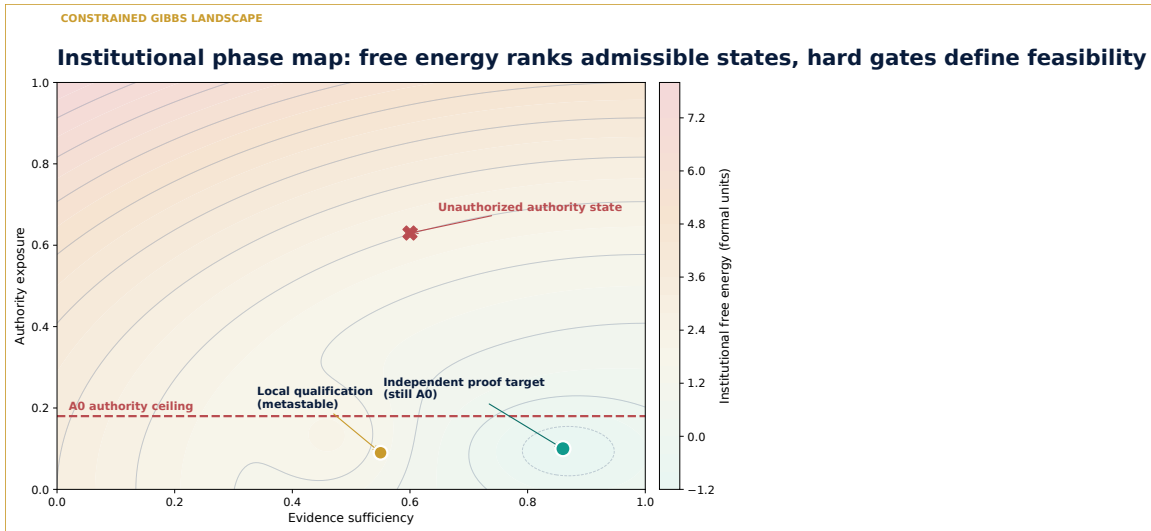


Figure 15.3.: Constrained phase map. Free energy ranks admissible states; hard proof and authority gates determine which states exist in the feasible set.

Table 15.1.: Decision meaning of the strategic-physical quantities.

Quantity	Institutional interpretation	What it cannot do
$H(x)$	Complete value-risk-evidence-authority score of one admissible microstate	Override a hard rule
τ	Declared uncertainty, volatility, and adversariality scale	Measure physical temperature
$H(q)$	Support entropy across admissible hypotheses or states	Certify a claim
$\mathcal{F}_\tau$	Risk-sensitive ranking of feasible state distributions	Grant proof, admission, or authority
$Z(\tau)$	Normalizer over admissible states	Create states excluded by the constitution

15.4 Evolutionary verifier ecology

One verifier can be honest while the verification institution remains fragile. Let $x \in [0, 1]$ denote the share of verifier capacity using a truthful strategy and let $q \in [0, 1]$ denote effective audit intensity. A stylized replicator dynamic is

$$\dot{x} = x(1 - x)(\Delta u_{T,C}(x, q)), \quad (15.6)$$

where $\Delta u_{T,C}$ is the payoff advantage of truthful verification over collusive or negligent verification. If the private manipulation benefit is b , the effective penalty is F , and detection probability is q , a basic deterrence condition is

$$qF \geq b. \quad (15.7)$$

A robust institution goes further: it reduces b through role separation and evidence custody, raises q through random audit and replayability, and raises the durable cost of manipulation through revocation, reputation, and access consequences.

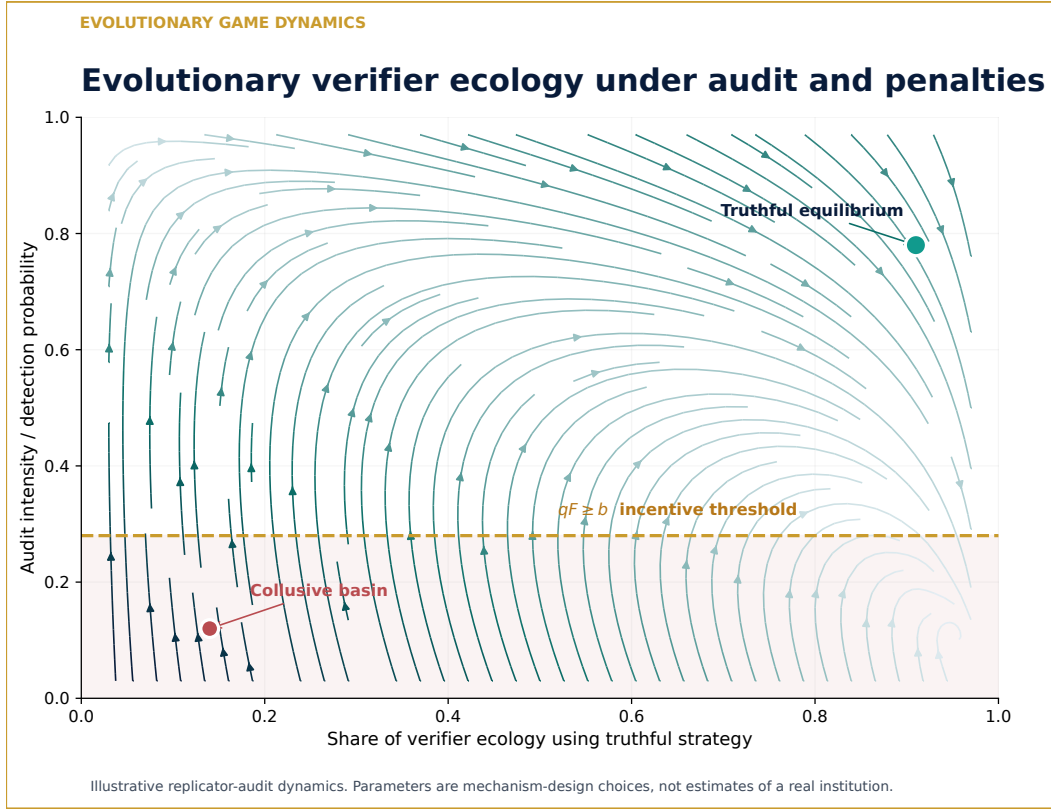


Figure 15.4.: Illustrative evolutionary verifier ecology. Proper incentives and audit intensity enlarge the basin of attraction of truthful verification.

The figure is not a fitted model of a real verifier market. It exposes a design requirement: truthful proof should be an equilibrium property of the mechanism rather than a personality assumption about one reviewer (Hofbauer and Sigmund, 1998; Başar and Olsder, 1999).

Proposition 15.2 (No-value purchase of proof). Let projected mission value affect the decision to commission proof, but not verifier compensation conditional on the proof outcome. If verifier reward is outcome-independent, evidence is independently controlled, and false attestation has expected cost at least equal to its private benefit, then higher projected value does not by itself make a false pass a best response.

15.5 Hamilton-Jacobi reachability under a hard authority envelope

Let x_t contain proof sufficiency, drift, stakeholder exposure, operational reversibility, and authority state. Let u_t be a proposed institutional control. The safe set is

$$\mathcal{K} = \{x : h_j(x) \geq 0, j = 1, \dots, m\}, \quad (15.8)$$

where the h_j encode the mission constitution, proof gate, admission gate, and authority ceiling. A control-barrier condition requires an admissible u such that

$$\nabla h_j(x)^\top f(x, u) + \alpha_j(h_j(x)) \geq 0 \quad \text{for every active constraint } j. \quad (15.9)$$

The associated Hamilton-Jacobi reachability question is not “what is the shortest path to higher authority?” It is “which proof-improving trajectories remain inside the safe set while authority stays fixed until a distinct authorized transition?” (Fleming and Soner, 2006; Bertsekas, 2017).

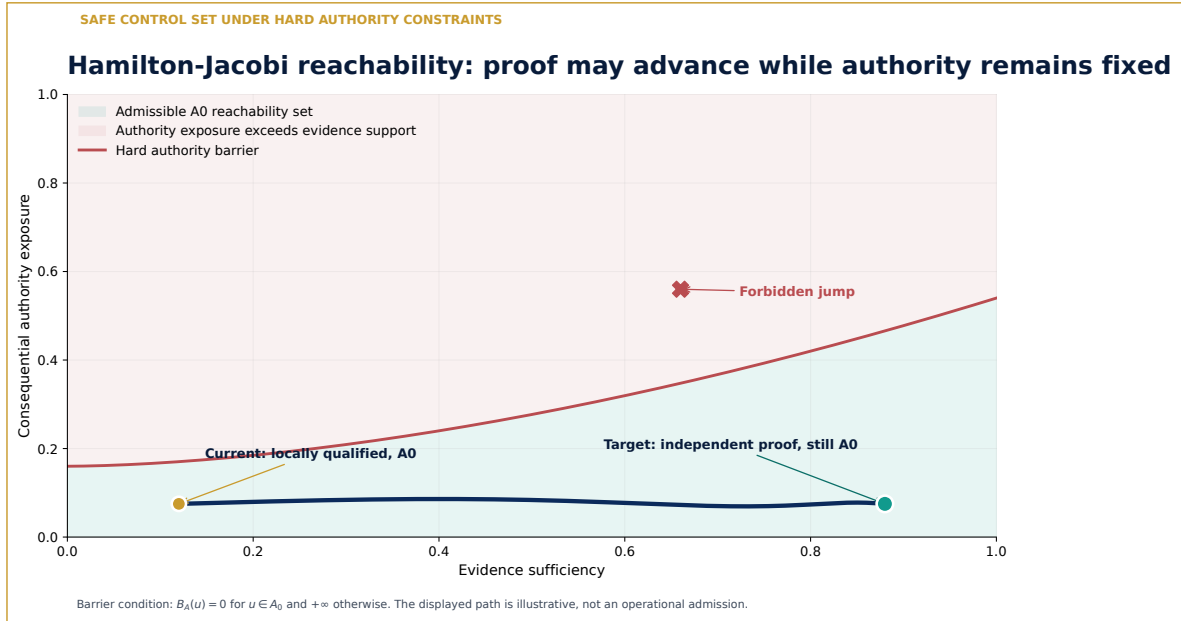


Figure 15.5.: Safe reachability. The exact candidate may move from local qualification toward independent proof without moving to a higher authority coordinate.

Principle 15.1 (Authority non-covariance). Evidence and capability trajectories are not required to covary with authority. The default safe policy is to hold $A = A_0$ while C and P evolve, and to treat any change in D or A as a separately signed institutional event.

15.6 Large-deviation refusal boundary

Rare but catastrophic governance failures should not disappear inside average performance. Let Y_n be an empirical institutional loss statistic over n fresh cases. Under regularity conditions, large-deviation theory characterizes tail probabilities through a rate function I :

$$\Pr(Y_n \in F) \asymp \exp\{-n \inf_{y \in F} I(y)\}. \quad (15.10)$$

For a forbidden set F containing unauthorized actions, critical stakeholder harm, or evidence-chain failure, the institution should precommit a minimum acceptable rate exponent or a direct zero-tolerance gate. This is stricter than comparing means. It asks whether dangerous states vanish rapidly enough with evidence or remain structurally reachable (Dembo and Zeitouni, 1998).

A complementary tail-risk measure is conditional value at risk:

$$\text{CVaR}_\beta(L) = \min_{\eta} \left\{ \eta + \frac{1}{1-\beta} \mathbb{E}[(L - \eta)_+] \right\}. \quad (15.11)$$

The proof contract may require both a positive lower confidence bound on complete-denominator Alpha and a bounded CVaR_β for false interventions, reviewer burden, and continuity loss (Rockafellar and Uryasev, 2000).

15.7 Joint eligibility functional

The mathematical layers can be combined without collapsing the institutional coordinates. Define

$$\mathcal{J} = \underbrace{1\{\underline{\alpha} > h\}}_{\text{value}} \underbrace{1\{P = P_{\text{independent}}\}}_{\text{proof}} \underbrace{1\{\Delta \mathcal{F}_\tau < 0\}}_{\text{risk-sensitive stability}} \underbrace{1\{\text{CVaR}_\beta(L) \leq \ell_{\max}\}}_{\text{tail control}} \prod_{k=1}^K \underbrace{g_k}_{\text{hard gates}}. \quad (15.12)$$

If $\mathcal{J} = 0$, admission is ineligible. If $\mathcal{J} = 1$, admission is merely eligible for a separate accountable decision. No scalar objective, posterior probability, free-energy minimum, or game equilibrium writes the admission receipt.

Table 15.2.: Board-facing refusal logic.

Condition	Interpretation	Required action
Local Alpha positive; proof open	Investment thesis exists, proof claim does not	Commission fresh proof
Independent proof passes; admission open	Superiority established for exact scope	Admission body decides
Admission passes; authority A0	Deployment state exists, action right does not	Keep A0 or issue separate authority decision
Tail-risk gate fails	Mean value is insufficient to justify exposure	Refuse or redesign
Role-independence gate fails	Evidence cannot support the claimed proof grade	Stop and reconstitute roles
Drift or identity binding fails	The evaluated object or environment changed	Re-freeze and requalify

15.8 The institutional conclusion

The frontier mathematical synthesis yields a disciplined asymmetry:

$$\begin{aligned} &\text{Value may justify proof expenditure,} \\ &\text{but value cannot purchase proof, admission, or authority.} \end{aligned} \quad (15.13)$$

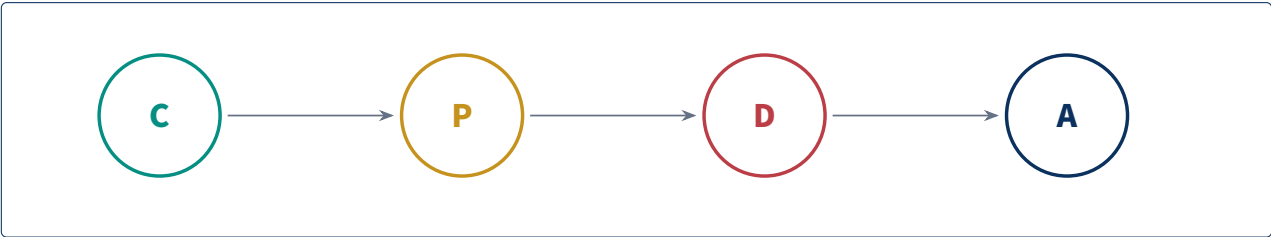
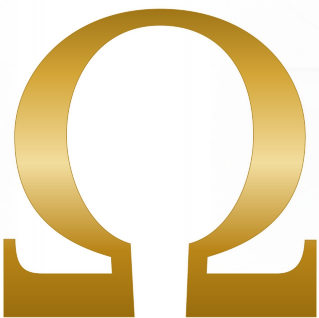
The exact candidate is therefore decision-relevant in the strongest legitimate sense available to the present evidence: it is valuable enough to justify a high-quality independent proof programme, yet bounded enough to remain A0 while that proof is open.

Frontier decision rule

Proceed only with the exact frozen candidate, fresh post-freeze evidence, authenticated comparators, independent custody, outcome-independent verification, precommitted value and tail-risk criteria, and separately constituted admission and authority decisions. Any missing element returns the state to local qualification or refusal.

Independent Proof, Admission, and the Next Institutional State

The verifier-ready proof contract, accountable admission,
bounded authority, limitations, and the next decision.



VALUE MAY JUSTIFY PROOF; VALUE NEVER PURCHASES PROOF, ADMISSION, OR AUTHORITY.

16 From Proof Candidate to Specialist ASI

16.1 The 90-day proof programme

The simulation emits a proof plan whose current status is `READY_FOR_INDEPENDENT_PROOF__NOT_PROOF`. The candidate digest is fixed, the duration is 90 days, and authority remains A0.

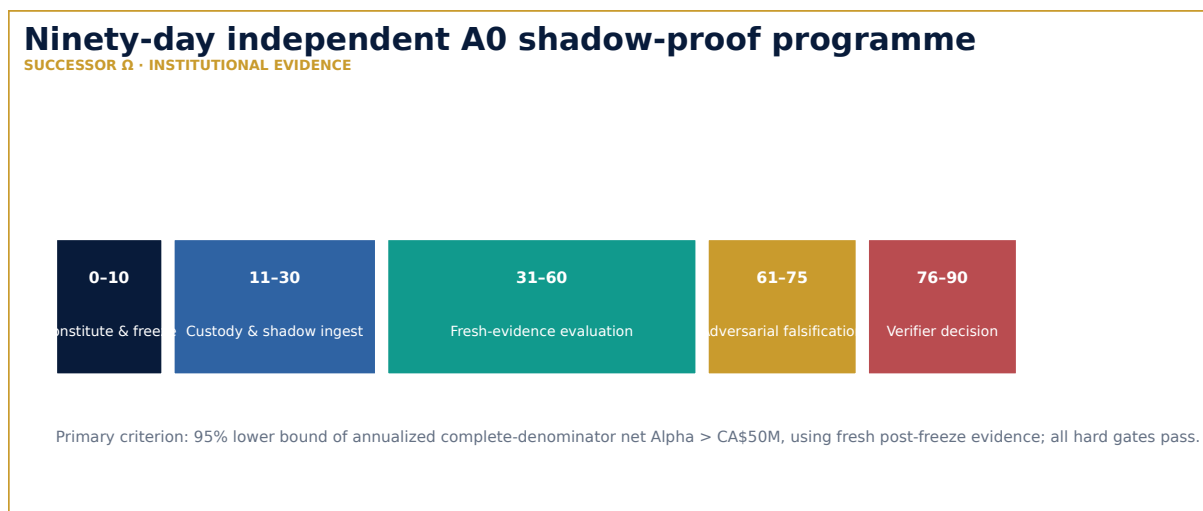


Figure 16.1.: Proposed 90-day independent A0 shadow-proof programme.

16.2 Precommitted primary criterion

The primary claim is:

The 95% lower bound of annualized complete-denominator net Alpha exceeds CA\$ 50M.
The comparator is the stronger of the authenticated incumbent and frozen Beta; only fresh post-freeze evidence is admissible, and every hard gate must pass.

The proof must bind the exact candidate, evidence window, comparator provenance, metric, unit, deductions, uncertainty method, threshold, and role credentials. No candidate retraining or threshold tuning may occur on the proof data.

16.3 Hard gates

1. zero unauthorized external effects;
2. zero consequential action without human approval;
3. candidate bytes and weights remain frozen;
4. no training or threshold tuning on fresh holdout evidence;
5. complete cost, risk, human burden, imitation, and drift deductions included;
6. every failure and abstention retained;
7. current comparator provenance authenticated;
8. verifier credential current and independently validated.

16.4 What a successful proof would and would not establish

A successful proof would establish mission-specific superiority for the exact frozen candidate under the stated proof contract. It would justify the label *independently proven Specialist ASI* only for that mission, evidence regime, candidate identity, and comparator constitution. It would not establish general intelligence, universal superiority, production admission, or a right to act.

Designation rule

The Specialist ASI designation is earned by fresh independent proof. A production admission is earned by a separate accountable decision. Consequential authority is earned by a further bounded-authority decision. No designation automatically grants the next state.

17 Institutional Decision, Limitations, and Next Stage

17.1 Decision synthesis

The local candidate is small, exact, and mission-specific; the institutional value case is large because the fictional portfolio is large. This asymmetry is precisely why governance matters. A small policy error can scale into a large exposure, while a small verified advantage can scale into a valuable institutional asset.

The evidence supports investment in proof, not investment in authority. The central and conservative scenario values are both above the hurdle. The Monte Carlo model assigns substantial mass above the hurdle. Most stress ecologies remain positive, but combined conservative conditions fall below the hurdle and extreme ecologies require halt. The learned policy is usefully active, yet symbolic rules downgrade proposals and A0 blocks external action.

17.2 Limitations

1. **Synthetic local evidence.** No real supplier payment or cash recovery is observed.
2. **Fictional portfolio.** The CA\$ 48.8B envelope and one million decision opportunities are scenario assumptions.
3. **Comparator provenance.** The supplied Beta actions are frozen but not independently authenticated.
4. **Independence.** Candidate formation and frontier simulation share an artifact lineage; this is not organizationally independent proof.
5. **Transfer model.** Retention distributions and stress parameters are transparent assumptions, not empirically fitted operational laws.
6. **Dependence structure.** Monte Carlo factors are sampled independently; real factors may be correlated.
7. **Model scope.** The policy is a bounded single-step decision model, not a general planner.
8. **Authority scope.** A0 operation cannot establish the safety of higher-authority action routes.
9. **Version lineage.** The study remains bound to R2, while the current R3 operational successor resolves the Console route; a future independent proof must name exactly which release it evaluates.
10. **Hash-chain limits.** The Chronicle is tamper-evident relative to its presented head but not globally unique or externally time-anchored.

17.3 Go / no-go decision

Table 17.1.: Decision matrix for the next institutional stage.

Decision	Recommendation	Reason
Fund independent 90-day A0 shadow proof	PASS	Conservative value case exceeds hurdle
Declare real-world Alpha	FAIL	No operational measurement
Declare Specialist ASI now	FAIL	Independent proof not completed
Admit to customer production	FAIL	No proof receipt or admission decision
Grant authority above A0	FAIL	No action-path evidence or authorization
Correct Operator Console in new release	PASS	Known route defect; version separately

17.4 Conclusion

The study demonstrates a disciplined path from model capability to institutional intelligence. The exact trained object preserves mission, comparator, evidence, model, programmes, tests, Alpha, authority, and Chronicle. The local candidate produces positive comparator-relative value. A transparent portfolio model shows that the signal could be economically significant after aggressive deductions. Distributional and adversarial analysis identifies where the thesis remains strong and where the system must halt. The proof plan converts those results into a falsifiable next stage.

The central institutional achievement is restraint: the simulation generates a powerful value thesis without manufacturing a crown. The candidate is ready to be challenged independently. Until it survives that challenge, it remains a frontier Specialist ASI proof candidate under A0 authority.

A Formulae and Assumption Ledger

A.1 Core equations

$$B^* = \max\{V_I, V_\beta\}, \quad (\text{A.1})$$

$$\alpha_{\text{local}}^* = V_\Omega - B^*, \quad (\text{A.2})$$

$$G = \alpha_{\text{local}}^* \frac{N}{n}, \quad (\text{A.3})$$

$$\alpha_{\text{portfolio}} = \rho G - (C + R + H + I + D), \quad (\text{A.4})$$

$$\alpha_{\text{lower}} = \rho L_{\text{local}} \frac{N}{n} - (C + R + H + I + D). \quad (\text{A.5})$$

For this simulation: $N = 1,000,000$, $n = 200$, $\rho = 0.72$, and $C + R + H + I + D = \text{CA\$ } 103\text{M}$.

A.2 Truth-preserving accounting

The following values must not be conflated:

- local synthetic holdout spread: established within the local protocol;
- portfolio scenario Alpha: established as a calculation under disclosed assumptions;
- realized operational Alpha: unmeasured;
- credited real-world Alpha: CA\$ 0.

B Monte Carlo Distribution Specification

Table B.1.: Disclosed uncertainty distributions.

Variable	Distribution	Interpretation
Technical transfer T	Beta(24, 4)	Model and process transfer
Coverage K	Beta(38, 2)	Fraction of eligible decisions with usable evidence
Adoption A	Beta(27, 3)	Fraction of eligible recommendations operationally reviewed
Adversarial retention E	Beta(18, 4)	Residual value after strategic adaptation
Evidence quality Q	Beta(29, 3)	Completeness and reliability of evidence
Complete cost C	Triangular(22, 32, 45) M	Implementation and operating cost
Risk reserve R	Triangular(16, 29, 48) M	Downside and continuity reserve
Human burden H	Triangular(8, 14, 24) M	Review and escalation burden
Imitation reserve I	Triangular(4, 11, 20) M	Comparator improvement reserve
Drift reserve D	Triangular(7, 17, 32) M	Monitoring and requalification reserve

The simulation assumes independence among these factors. A later proof should model or bound correlations where operational evidence supports them.

C Exact-Model Probe Atlas Summary

Table C.1.: Exact-model probe summary.

Metric	Result
Structured probes	96
Neural proposals changed by symbolic rules	19
Bounded OBSERVE outputs	72
Bounded RECOMMEND outputs	22
Bounded REVERSIBLE_EXPERIMENT outputs	2
External effects	0
Effective authority	A0

The complete probe-level features, scores, neural actions, symbolic reasons, and bounded outputs are included in ‘EXACT_MODEL_PROBE_ATLAS.json’.

D Independent Proof Contract

Table D.1.: Proof contract summary.

Field	Precommitted value
Candidate	SHA-256 142706131d9b017e0c20c5f09a127cf66d700f0a7989bd9f4f302d29afddb915
Duration	90 days
Authority	A0
Evidence	Fresh, post-freeze, independently controlled
Comparator	Stronger of authenticated incumbent and frozen Beta
Primary metric	Annualized complete-denominator net Alpha
Criterion	95% lower bound greater than CA\$ 50M
Training on proof evidence	Prohibited
Admission	Separate decision after proof
Authority escalation	Separate decision after admission

E Artifact Inventory and Digests

Table E.1.: Principal released artifacts.

Artifact	Identity or purpose
Research-bound R2 Buyer Package	f7747d176def4594b1f65f642a54c8ba49d8396747f736eaced1ab700dd27e4e
Current operational R3 Buyer Package	3f15b848306e96d2825fd2a5ec68b112dd13b49806685e93986259e2498ac67a
Trained candidate	142706131d9b017e0c20c5f09a127cf66d700f0a7989bd9f4f302d29afddb915
Alpha ledger	17c6746eb82ca12ff02667f31598abc58d079b41c316e8e1d168b639d47dba9c
Proof plan	bb7eaa4c8bf31a8121aac18ca1106a6c5f4a760888052f2dfe0d4f2d61d0e67b
Chronicle head	0f22eef59acdd5b1c544ba4a337719d046964e52670e053c92ea0ab48ee3bf19
Simulation report	0f9b2ce9148ba9b9d1c70c0c3db687acb784f832887c3cfd28380100aef3de93

F Reporting and Reproducibility Checklist

Table F.1.: Artifact-grounded reporting checklist.

Requirement	Status	Location
Exact artifact identities	PASS	Chapter 4, Appendix E
Data partitions and comparator logic	PASS	Chapters 6–7
Candidate selection and freeze	PASS	Chapter 6
Metrics and uncertainty	PASS	Chapters 7–9
Negative and stress results	PASS	Chapter 10
Authority and external effects	PASS	Chapters 2, 5, 11
Known defects	PASS	Chapter 11
Independent-proof boundary	PASS	Chapters 2, 13
Source and figure regeneration	PASS	Source package
Deterministic paper rebuild	PASS	Validation report

G Conflict, Independence, and Publication Disclosure

This is an artifact-grounded technical report authored within the same project lineage as the product and simulation. It is not peer review, an independent audit, external certification, legal advice, or a customer outcome study. The institution, persona, portfolio, and all portfolio monetary results are fictional. The report's purpose is to make the evidence, assumptions, value thesis, limitations, and proof obligations inspectable.

Material changes to the candidate, comparator, mission, data-generating process, metric, deductions, or authority envelope create a new claim and require a new evidence record. A reader should treat the released digests as the identity of the evaluated objects, not as a guarantee of their external provenance or truth.

References

- Alshiekh, Mohammed, Roderick Bloem, Ruediger Ehlers, Bettina Koenighofer, Scott Niekum, and Ufuk Topcu (2018). “Safe Reinforcement Learning via Shielding”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. DOI: [10.1609/aaai.v32i1.11797](https://doi.org/10.1609/aaai.v32i1.11797).
- Arjovsky, Martin, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz (2019). “Invariant Risk Minimization”. In: *arXiv preprint arXiv:1907.02893*. DOI: [10.48550/arXiv.1907.02893](https://doi.org/10.48550/arXiv.1907.02893).
- Başar, Tamer and Geert Jan Olsder (1999). *Dynamic Noncooperative Game Theory*. 2nd ed. SIAM.
- Bertsekas, Dimitri P. (2017). *Dynamic Programming and Optimal Control*. 4th ed. Athena Scientific.
- Besold, Tarek R., Artur d’Avila Garcez, Sebastian Bader, Howard Bowman, Pedro Domingos, Pascal Hitzler, Kai-Uwe Kuehnberger, Luis C. Lamb, Daniel Lowd, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha (2017). “Neural-Symbolic Learning and Reasoning: A Survey and Interpretation”. In: *arXiv preprint arXiv:1711.03902*. DOI: [10.48550/arXiv.1711.03902](https://doi.org/10.48550/arXiv.1711.03902).
- Breck, Eric, Shanqing Cai, Eric Nielsen, Michael Salib, and D. Sculley (2017). “The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction”. In: *2017 IEEE International Conference on Big Data*. IEEE, pp. 1123–1132. DOI: [10.1109/BigData.2017.8258038](https://doi.org/10.1109/BigData.2017.8258038).
- Brockman, Greg, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba (2016). “OpenAI Gym”. In: *arXiv preprint arXiv:1606.01540*. DOI: [10.48550/arXiv.1606.01540](https://doi.org/10.48550/arXiv.1606.01540).
- Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory*. 2nd ed. Wiley-Interscience.
- Crooks, Gavin E. (1999). “Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences”. In: *Physical Review E* 60.3, pp. 2721–2726. DOI: [10.1103/PhysRevE.60.2721](https://doi.org/10.1103/PhysRevE.60.2721).
- Dalrymple, David, Joar Skalse, Yoshua Bengio, Stuart Russell, Max Tegmark, Sanjit Seshia, Steve Omohundro, Christian Szegedy, Ben Goldhaber, Nora Ammann, Alessandro Abate, Joseph Y. Halpern, Clark Barrett, Ding Zhao, Zhi-Xuan Tan, Jeannette M. Wing, and Joshua B. Tenenbaum (2024). “Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems”. In: *arXiv preprint arXiv:2405.06624*. DOI: [10.48550/arXiv.2405.06624](https://doi.org/10.48550/arXiv.2405.06624).
- Dembo, Amir and Ofer Zeitouni (1998). *Large Deviations Techniques and Applications*. 2nd ed. Springer.
- Dwork, Cynthia, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth (2014). “Preserving Statistical Validity in Adaptive Data Analysis”. In: *arXiv preprint arXiv:1411.2664*. DOI: [10.48550/arXiv.1411.2664](https://doi.org/10.48550/arXiv.1411.2664).
- (2015). “Generalization in Adaptive Data Analysis and Holdout Reuse”. In: *Advances in Neural Information Processing Systems*. Vol. 28.
- Fleming, Wendell H. and H. Mete Soner (2006). *Controlled Markov Processes and Viscosity Solutions*. 2nd ed. Springer.
- Gebri, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford (2021). “Datasheets for Datasets”. In: *Communications of the ACM* 64.12, pp. 86–92. DOI: [10.1145/3458723](https://doi.org/10.1145/3458723).
- Gulrajani, Ishaan and David Lopez-Paz (2021). “In Search of Lost Domain Generalization”. In: *International Conference on Learning Representations*.
- Haber, Stuart and W. Scott Stornetta (1991). “How to Time-Stamp a Digital Document”. In: *Journal of Cryptology* 3, pp. 99–111. DOI: [10.1007/BF00196791](https://doi.org/10.1007/BF00196791).

- Hairer, Ernst, Christian Lubich, and Gerhard Wanner (2006). *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. 2nd ed. Springer. DOI: [10.1007/3-540-30666-8](https://doi.org/10.1007/3-540-30666-8).
- Hofbauer, Josef and Karl Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.
- Howard, Steven R., Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon (2021). “Time-Uniform, Nonparametric, Nonasymptotic Confidence Sequences”. In: *The Annals of Statistics* 49.2, pp. 1055–1080. DOI: [10.1214/20-AOS1991](https://doi.org/10.1214/20-AOS1991).
- Jarzynski, Christopher (1997). “Nonequilibrium Equality for Free Energy Differences”. In: *Physical Review Letters* 78.14, pp. 2690–2693. DOI: [10.1103/PhysRevLett.78.2690](https://doi.org/10.1103/PhysRevLett.78.2690).
- Jaynes, Edwin T. (1957). “Information Theory and Statistical Mechanics”. In: *Physical Review* 106.4, pp. 620–630. DOI: [10.1103/PhysRev.106.620](https://doi.org/10.1103/PhysRev.106.620).
- Kapoor, Sayash, Emily M. Cantrell, Kenny Peng, Thanh Hien Pham, Christopher A. Bail, Odd Erik Gundersen, Jake M. Hofman, Jessica Hullman, Michael A. Lones, Momin M. Malik, Priyanka Nanayakkara, Russell A. Poldrack, Inioluwa Deborah Raji, Michael Roberts, Matthew J. Salganik, Marta Serra-Garcia, Brandon M. Stewart, Gilles Vandewiele, and Arvind Narayanan (2024). “REFORMS: Consensus-Based Recommendations for Machine-Learning-Based Science”. In: *Science Advances* 10.18, eadk3452. DOI: [10.1126/sciadv.adk3452](https://doi.org/10.1126/sciadv.adk3452).
- Kapoor, Sayash and Arvind Narayanan (2023). “Leakage and the Reproducibility Crisis in Machine-Learning-Based Science”. In: *Patterns* 4.9, p. 100804. DOI: [10.1016/j.patter.2023.100804](https://doi.org/10.1016/j.patter.2023.100804).
- Kappen, Hilbert J. (2005). “Path Integrals and Symmetry Breaking for Optimal Control Theory”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2005.11, P11011. DOI: [10.1088/1742-5468/2005/11/P11011](https://doi.org/10.1088/1742-5468/2005/11/P11011).
- Kullback, Solomon and Richard A. Leibler (1951). “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86. DOI: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- Li, Lihong, Wei Chu, John Langford, and Robert E. Schapire (2010). “A Contextual-Bandit Approach to Personalized News Article Recommendation”. In: *Proceedings of the 19th International Conference on World Wide Web*. ACM, pp. 661–670. DOI: [10.1145/1772690.1772758](https://doi.org/10.1145/1772690.1772758).
- Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru (2019). “Model Cards for Model Reporting”. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, pp. 220–229. DOI: [10.1145/3287560.3287596](https://doi.org/10.1145/3287560.3287596).
- Monderer, Dov and Lloyd S. Shapley (1996). “Potential Games”. In: *Games and Economic Behavior* 14.1, pp. 124–143. DOI: [10.1006/game.1996.0044](https://doi.org/10.1006/game.1996.0044).
- Myerson, Roger B. (1981). “Optimal Auction Design”. In: *Mathematics of Operations Research* 6.1, pp. 58–73. DOI: [10.1287/moor.6.1.58](https://doi.org/10.1287/moor.6.1.58).
- Nash, John F. (1950). “Equilibrium Points in N-Person Games”. In: *Proceedings of the National Academy of Sciences* 36.1, pp. 48–49. DOI: [10.1073/pnas.36.1.48](https://doi.org/10.1073/pnas.36.1.48).
- National Institute of Standards and Technology (2015). *Secure Hash Standard (SHS)*. Tech. rep. FIPS PUB 180-4. U.S. Department of Commerce. DOI: [10.6028/NIST.FIPS.180-4](https://doi.org/10.6028/NIST.FIPS.180-4).
- (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Tech. rep. NIST AI 100-1. U.S. Department of Commerce. DOI: [10.6028/NIST.AI.100-1](https://doi.org/10.6028/NIST.AI.100-1).
- Necula, George C. (1997). “Proof-Carrying Code”. In: *Proceedings of the 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. ACM, pp. 106–119. DOI: [10.1145/263699.263712](https://doi.org/10.1145/263699.263712).
- Ortega, Pedro A. and Daniel A. Braun (2013). “Thermodynamics as a Theory of Decision-Making with Information-Processing Costs”. In: *Proceedings of the Royal Society A* 469.2153, p. 20120683. DOI: [10.1098/rspa.2012.0683](https://doi.org/10.1098/rspa.2012.0683).
- Pathria, R. K. and Paul D. Beale (2011). *Statistical Mechanics*. 3rd ed. Elsevier.
- Phan, Dung T., Radu Grosu, Nils Jansen, Nicola Paoletti, Scott A. Smolka, and Scott D. Stoller (2020). “Neural Simplex Architecture”. In: *arXiv preprint arXiv:1908.00528*. DOI: [10.48550/arXiv.1908.00528](https://doi.org/10.48550/arXiv.1908.00528).

- Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes (2020). “Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing”. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, pp. 33–44. DOI: [10.1145/3351095.3372873](https://doi.org/10.1145/3351095.3372873).
- Rockafellar, R. Tyrrell and Stanislav Uryasev (2000). “Optimization of Conditional Value-at-Risk”. In: *Journal of Risk* 2.3, pp. 21–41. DOI: [10.21314/JOR.2000.038](https://doi.org/10.21314/JOR.2000.038).
- Rundgren, Anders, Brian Jordan, and Samuel Erdtman (2020). *JSON Canonicalization Scheme (JCS)*. RFC 8785. DOI: [10.17487/RFC8785](https://doi.org/10.17487/RFC8785).
- Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola (2008). *Global Sensitivity Analysis: The Primer*. Wiley. DOI: [10.1002/9780470725184](https://doi.org/10.1002/9780470725184).
- Sculley, D., Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, and Dan Dennison (2015). “Hidden Technical Debt in Machine Learning Systems”. In: *Advances in Neural Information Processing Systems*. Vol. 28, pp. 2503–2511.
- Shannon, Claude E. (1948). “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3–4, pp. 379–423, 623–656.
- Sutton, Richard S. and Andrew G. Barto (2018). *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press.
- Todorov, Emanuel (2009). “Efficient Computation of Optimal Actions”. In: *Proceedings of the National Academy of Sciences* 106.28, pp. 11478–11483. DOI: [10.1073/pnas.0710743106](https://doi.org/10.1073/pnas.0710743106).
- Towers, Mark, Ariel Kwiatkowski, Jordan Terry, John U. Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, Rodrigo Perez-Vicente, Andrea Pierré, Sander Schulhoff, Jun Jet Tai, Hannah Tan, and Omar G. Younis (2024). “Gymnasium: A Standard Interface for Reinforcement Learning Environments”. In: *arXiv preprint arXiv:2407.17032*. Accepted at NeurIPS Datasets and Benchmarks 2025. DOI: [10.48550/arXiv.2407.17032](https://doi.org/10.48550/arXiv.2407.17032).

Intelligence de mission de pointe assujettie à la preuve

Équilibres stratégiques · thermodynamique de l'information · physique statistique
Contrôle hamiltonien de mission · énergie libre institutionnelle de style Gibbs



Une grande synthèse institutionnelle complète fondée sur les artéfacts reliant la formation exacte d'un candidat neuronal-symbolique, l'Alpha de mission relatif aux comparateurs, la souscription du portefeuille, le gain d'information, les jeux évolutifs de vérification, l'accessibilité de Hamilton-Jacobi, les frontières de refus fondées sur les grandes déviations et une gouvernance contrainte par l'énergie libre de Gibbs—tout en maintenant la preuve indépendante, l'admission responsable et l'autorité effective comme décisions institutionnelles distinctes.

GAME THEORY

Truthful equilibria · role separation

STATISTICAL PHYSICS

Admissible ensembles · uncertainty

HAMILTONIAN CONTROL

Reversible search · safe reachability

GIBBS FREE ENERGY

Value · risk · cost · admissibility

$$qF \geq b \quad | \quad \pi_\tau(\omega) \propto e^{-\mathcal{H}/\tau} \quad | \quad \mathcal{H}_p = \ell + \lambda^\top f + B_A \quad | \quad C \neq P \neq D \neq A$$

CANDIDAT DE PREUVE DE POINTE

A0 · 0 EFFET EXTERNE

FONDÉ SUR LES ARTÉFACTS

Vincent Boucher · Président, MONTREAL.AI & QUEBEC.AI

Grande monographie institutionnelle v8.0.0

2026-09-04

Note de décision institutionnelle

DÉCISION EN UNE PHRASE

Passer à une preuve fantôme A0 de 90 jours sous gouvernance indépendante pour le candidat exact figé; ne pas accorder d'admission en production, d'autorité conséquente, de valeur réalisée reconnue ou de désignation d'ASI spécialiste tant que des preuves nouvelles ne satisfont pas au contrat de preuve préengagé.

48,8 G \$ CA flux annuel fictif de paiements fournisseurs	1 000 000 occasions annuelles pondérées de décision	218,5 M \$ CA Alpha net central après toutes les déductions
70,5 M \$ CA scénario inférieur prudent après toutes les déductions	98,27% probabilité simulée de dépasser 50 M \$ CA	A0 / 0 autorité effective / effets externes

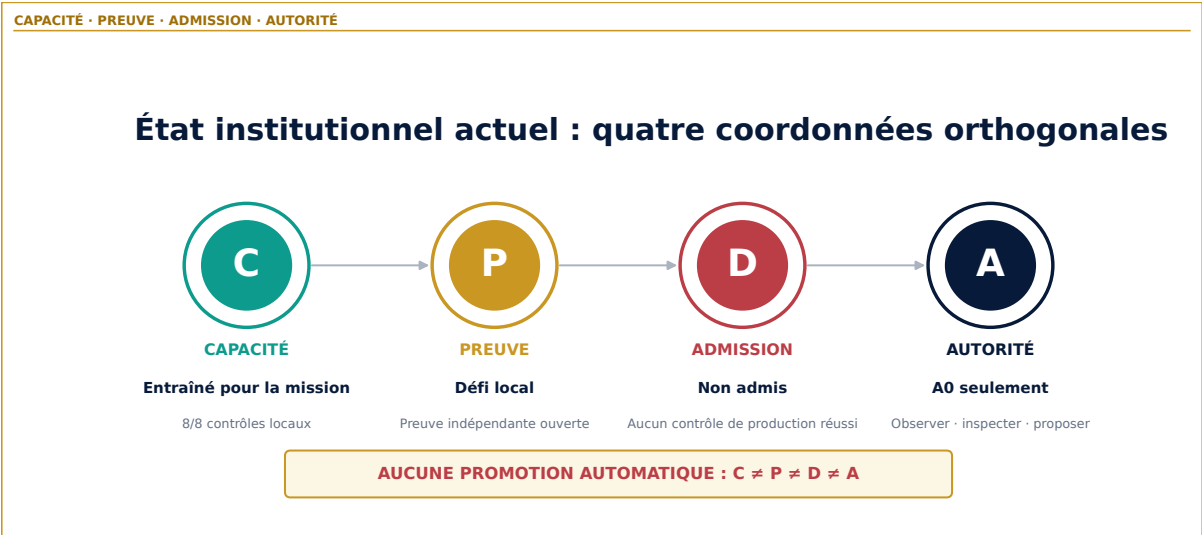
Interprétation pour le conseil

Le candidat est assez fort pour justifier un programme coûteux de preuve indépendante, mais pas assez pour justifier de le contourner. Le signal local survit à une perte de transfert explicite, aux déductions à dénominateur complet, à l'incertitude distributionnelle, à la sensibilité du transfert causal et à plusieurs écologies contradictoires. L'étude n'établit ni Alpha réel, ni preuve indépendante, ni aptitude à la production, ni droit d'agir.

TABLE 1. – Droits de décision et état probant actuel.

Coordonnée	État actuel	Décision
Capacité C	Candidat spécialiste neuronal-symbolique utile, entraîné pour la mission et figé	Poursuivre l'évaluation
Preuve P	Jeu de contrôle synthétique local réussi; aucune preuve indépendante sur données nouvelles	Commander une preuve indépendante
Admission D	Aucune admission dans la production du client	Refuser pour l'instant
Autorité A	A0 : observer, inspecter et proposer seulement; zéro effet externe	Maintenir A0

Architecture de l'état institutionnel



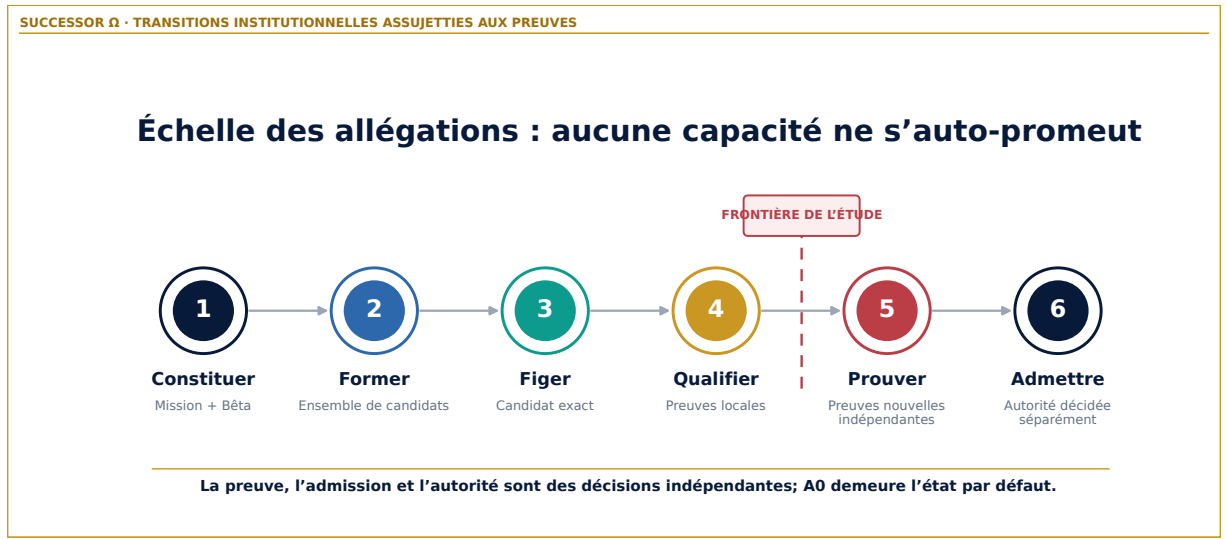
Capacité, preuve, admission et autorité constituent quatre coordonnées institutionnelles orthogonales.

Invariant directeur

Capacité ≠ Preuve ≠ Admission ≠ Autorité

Une preuve réussie ne modifie que la coordonnée de preuve. L'admission et l'autorité demeurent inchangées jusqu'à ce que des décisions distinctes soient constituées et consignées.

Échelle des preuves et frontière de refus



La présente étude s’arrête après la qualification locale. Les preuves nouvelles indépendantes constituent le prochain échelon ; l’admission et l’autorité viennent ensuite.

Aucune promotion automatique

La qualification locale ne peut émettre son propre reçu de preuve, s’admettre elle-même en production ou élargir son autorité. Chaque transition au-delà de la frontière de l’étude exige des preuves nouvelles, des rôles distincts, des droits décisionnels explicites et un dossier signé vérifiable.

Architecture mathématique de gouvernance



Un même candidat exact et figé est examiné au moyen de quatre lentilles analytiques distinctes.

Équilibre stratégique La vérification véridique devient compatible avec les incitations lorsque le coût attendu de la manipulation domine le gain privé : $qF \geq b$.	Ensemble d'états admissibles La distribution de Gibbs classe uniquement les états qui satisfont aux contrôles absolus de mission et d'autorité : $\pi_\tau(\omega) \propto e^{-H(\omega)/\tau}$.
Accessibilité sûre Le Hamiltonien de mission contraint optimise les trajectoires uniquement à l'intérieur de l'ensemble sûr A0 constitué.	Énergie libre institutionnelle Valeur, coût complet, risque extrême et entropie probante sont combinés sans transformer la valeur projetée en preuve ou en autorité.

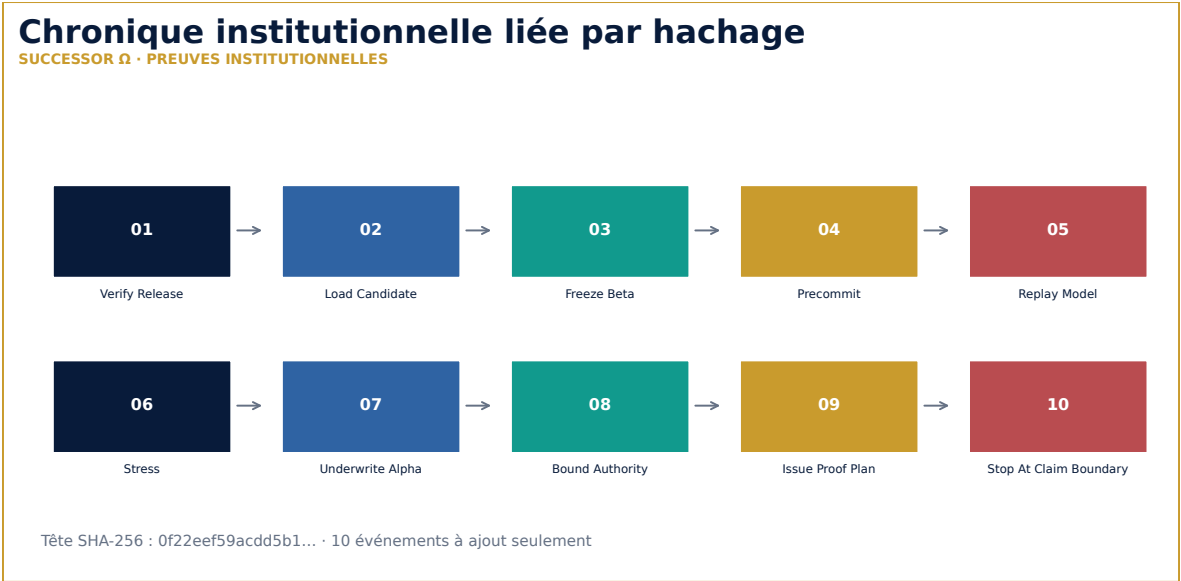
Charte des preuves

Établi par l'ensemble exact d'artéfacts	Non établi
— Identités exactes de la trousse acheteur et du candidat. — Rejeu déterministe des poids appris. — Alpha* positif du scénario local par rapport au comparateur fourni le plus fort. — Huit contrôles locaux réussis sur huit. — Objet scellé sous le contrôle du client et Chronique liée par hachage. — Souscription transparente d'un portefeuille fictif. — Plan exécutable de preuve indépendante.	— Économies réalisées ou Alpha réel. — Provenance authentifiée des comparateurs au-delà des dossiers fournis. — Preuve indépendante ou certification externe. — Admission en production chez le client. — Autorité supérieure à A0. — ASI spécialiste formellement démontrée. — Superintelligence générale ou universelle.
Règle d'interprétation non négociable	
Tout résultat monétaire supérieur à zéro dans le modèle de portefeuille constitue une <i>sortie fictive de souscription</i>. L'Alpha réel demeure non mesuré et reconnu à 0 \$ CA. Toute probabilité est conditionnelle aux distributions déclarées; elle ne constitue pas une probabilité empirique d'économies réalisées. Chaque résultat opérationnel demeure à A0 et ne produit aucun effet externe.	

Carte de lecture et filiation des artéfacts

TABLE 2. – Parcours de lecture conçus selon le rôle.

Lecteur	Parcours recommandé
Conseil / comité d’investissement	Note de décision → chapitres 1, 8, 11, 13, 14 et 16
Risque / audit / juridique	Charte des preuves → chapitres 2, 4, 10, 12, 14 et 15
Direction scientifique / ML	Chapitres 5 à 15 et annexes A à E
Vérificateur indépendant	Chapitres 4, 7, 10, 12, 14 et 15 et annexes C à G
Exploitant / responsable de mission	Chapitres 5, 6, 10 et 12 à 14



La filiation lie la trousse acheteur exacte, le candidat, les preuves, le registre Alpha, la Chronique et la demande de preuve.

Résumé exécutif

Cette étude fondée sur les artéfacts examine si une utilisatrice institutionnelle hautement technique peut transformer un candidat neuronal-symbolique exact, sous le contrôle du client, en candidat de preuve exploitable pour une décision sans confondre capacité locale, preuve indépendante ou autorité. La mission fictive, *Intégrité souveraine du capital fournisseurs* Ω , protège un flux annuel simulé de paiements fournisseurs de 48,8 G \$ CA dans 41 entités opérationnelles, auprès de 72 800 fournisseurs, dans 26 territoires et pour un million d'occasions annuelles pondérées de décision.

Le résultat local exact provient d'un jeu de contrôle scellé de 200 cas : le processus en place produit 882 697 \$ CA, le comparateur Bêta figé produit 1 237 761 \$ CA et le candidat figé produit 1 327 063 \$ CA. Bêta est donc le comparateur le plus fort, et l'Alpha* du scénario local atteint 89 302 \$ CA ; la borne inférieure normale de sensibilité est de 48 208 \$ CA. Le candidat sélectionné atteint 91,5% d'actions exactes, zéro expérience nuisible déclarée et huit contrôles locaux réussis sur huit.

La simulation de pointe lie la trousse acheteur exacte, l'objet entraîné, la demande de preuve et les preuves du moteur A0 ; rejoue les poids appris exacts sur 96 sondes structurées ; consigne 19 déclassements symboliques ; transpose le signal local à un million d'occasions fictives de décision ; applique une réduction de transfert de 28% et 103 M \$ CA de déductions à dénominateur complet ; exécute 250 000 tirages distributionnels ; évalue 13 écologies de stress contradictoires ; écrit un registre Alpha résistant à l'altération ainsi qu'une Chronique de dix événements ; puis produit un plan de preuve fantôme indépendante A0 de 90 jours.

L'Alpha net central après toutes les déductions est de 218,4872 M \$ CA. Le calcul prudent qui commence par la borne inférieure locale atteint 70,5488 M \$ CA, soit une marge de 20,5488 M \$ CA au-dessus du seuil préengagé de 50 M \$ CA. Dans les écologies extrêmes d'équipe rouge et d'arrêt de sûreté, le système s'arrête et reconnaît un Alpha de scénario nul plutôt que de masquer l'échec par une moyenne.

La conclusion la plus forte et exacte est que l'objet constitue un *candidat de preuve d'ASI spécialiste de pointe* : utile, entraîné, qualifié localement, portable, borné et suffisamment précieux pour justifier une preuve indépendante. L'étude n'établit ni ASI spécialiste, ni économies réalisées, ni admission en production, ni autorité supérieure à A0.

Questions de recherche

1. Des preuves locales exactes peuvent-elles devenir une thèse transparente de souscription à l'échelle d'un portefeuille sans être présentées comme valeur réalisée ?
2. La politique apprise demeure-t-elle utile lorsqu'elle est subordonnée à des contraintes symboliques déterministes et à une autorité A0 ?
3. Dans quelle mesure la thèse résiste-t-elle aux pertes de transfert, aux coûts complets, à la dérive, à l'adaptation contradictoire, à la dégradation des preuves et à l'amélioration du comparateur ?

4. Quelles preuves et séparations de rôles sont requises avant de désigner le candidat comme ASI spécialiste démontrée indépendamment ?
5. Les incitations de théorie des jeux et les modèles statistico-physiques contraints peuvent-ils rendre les frontières de preuve et d'autorité plus utiles à la décision sans gonfler l'allégation empirique ?

Table des matières

Note de décision institutionnelle	i
Architecture de l'état institutionnel	ii
Échelle des preuves et frontière de refus	iii
Architecture mathématique de gouvernance	iv
Charte des preuves	v
Carte de lecture et filiation des artéfacts	vi
Résumé exécutif	vii
Fondements de l'intelligence de mission assujettie à la preuve	1
1. Thèse institutionnelle	2
1.1. L'unité de valeur n'est pas le modèle	2
1.2. Thèse de décision	2
1.3. Questions de recherche et résultats falsifiables	3
2. Système d'allégations et définitions opérationnelles	4
2.1. Échelle des preuves	4
2.2. Définitions	4
2.3. Matrice des allégations autorisées	6
3. Travaux connexes et filiation conceptuelle	7
3.1. Systèmes neuronaux-symboliques et assurance à l'exécution	7
3.2. Évaluation comparative, intégrité du jeu de contrôle et changement de domaine	7
3.3. Risque, incertitude et preuves de gouvernance	7
3.4. Objets porteurs de preuves et historique résistant à l'altération	8
4. Filiation exacte des artéfacts et reproductibilité	9
4.1. Intrants liés	9
4.2. Contrôles de reproductibilité	10
4.3. Sémantique de la Chronique	10
5. Constitution de mission et enveloppe de portefeuille	11
5.1. Institution fictive et mission	11
5.2. Contraintes non négociables	11
5.3. Pourquoi l'intégrité du capital fournisseurs est une mission exigeante	12

6. Architecture institutionnelle neuronale-symbolique	13
6.1. Boucle de fonctionnement	13
6.2. Le spécialiste est le composite	13
6.3. Capacité, preuve, admission et autorité	14
7. Conception expérimentale et candidat exact	15
7.1. Filiation des preuves locales	15
7.2. Rejeu du modèle exact	15
7.3. Contrôles absolus locaux	16
8. Qualification locale et formation du candidat	17
8.1. Résultat local exact	17
8.2. Pourquoi le comparateur le plus fort compte	18
8.3. Pourquoi le succès local demeure insuffisant	18
Alpha, incertitude et robustesse contradictoire	19
9. Souscription du portefeuille à dénominateur complet	20
9.1. Modèle d'échelle	20
9.2. Rétention de transfert et dénominateur complet	20
9.3. Scénario inférieur prudent	21
9.4. Valeur économique de l'information	22
10. Incertitude distributionnelle et transfert causal	23
10.1. Modèle Monte-Carlo	23
10.2. Résultats distributionnels	24
10.3. Sensibilité du transfert causal	24
11. Écologie institutionnelle contradictoire	26
11.1. Conception	26
11.2. Interprétation	27
11.3. Tableau complet des stress	27
12. Moteur A0, bouclier symbolique et Chronique	28
12.1. Politique exacte dans des contraintes constitutionnelles	28
12.2. Exploitation A0 authentifiée	28
12.3. Successeur opérationnel R3 et filiation immuable de l'étude	29
Équilibres stratégiques et thermodynamique institutionnelle	30
13. Gouvernance institutionnelle et droits décisionnels	31
13.1. Séparation des rôles	31
13.2. Objets décisionnels	31
13.3. Contrôles du conseil	32
14. Gouvernance stratégique et physique et conception du mécanisme	33
14.1. Finalité et interprétation	33

14.2. Le jeu institutionnel de la preuve	34
14.2.1. Représentation comme jeu potentiel	35
14.3. Microétats institutionnels et ensemble de Gibbs	36
14.4. Formation hamiltonienne avant la frontière de figement	37
14.5. Température de gouvernance et transitions de phase	38
14.6. Relier l'Alpha de mission au budget d'énergie libre	39
14.7. Protocole d'exploitation stratégique et physique	40
14.8. Limites du formalisme stratégique et physique	41
15. Synthèse de pointe : équilibres stratégiques, thermodynamique de l'information et accessibilité sûre	42
15.1. Objet : un calcul décisionnel, non une métaphore	42
15.2. Les preuves comme gain d'information	43
15.3. Valeur sensible au risque et énergie libre de style Gibbs	44
15.4. Écologie évolutive des vérificateurs	45
15.5. Accessibilité de Hamilton-Jacobi sous une enveloppe d'autorité dure	46
15.6. Frontière de refus fondée sur les grandes déviations	47
15.7. Fonctionnelle conjointe d'admissibilité	48
15.8. Conclusion institutionnelle	48
Preuve indépendante, admission et prochain état institutionnel	49
16. Du candidat de preuve à l'ASI spécialiste	50
16.1. Programme de preuve de 90 jours	50
16.2. Critère principal préengagé	50
16.3. Contrôles absolus	50
16.4. Ce qu'une preuve réussie établirait — et n'établirait pas	51
17. Décision institutionnelle, limites et prochaine étape	52
17.1. Synthèse décisionnelle	52
17.2. Limites	52
17.3. Décision autorisé / refusé	53
17.4. Conclusion	53
A. Formules et registre des hypothèses	54
A.1. Équations principales	54
A.2. Comptabilité préservant la vérité	54
B. Spécification des distributions Monte-Carlo	55
C. Résumé de l'atlas des sondes du modèle exact	56
D. Contrat de preuve indépendante	57
E. Inventaire des artéfacts et empreintes	58
F. Liste de vérification de la communication et de la reproductibilité	59
G. Divulgence relative aux conflits, à l'indépendance et à la publication	60

Table des figures

2.1. Échelle des allégations. La qualification locale ne constitue ni preuve indépendante, ni admission, ni autorité.	4
4.1. Filiation des artéfacts, des intrants exacts au dossier de preuve indépendante.	9
4.2. Chronique de simulation à dix événements.	10
6.1. Pipeline de valeur institutionnelle représenté dans la trousse exacte de simulation.	13
6.2. Séparation de la capacité, de la preuve, de l'admission et de l'autorité dans la simulation.	14
7.1. Répartition des actions bornées lors du rejou du modèle exact.	16
8.1. Registre exact de valeur locale sur le jeu de contrôle scellé de 200 cas.	17
9.1. Souscription à dénominateur complet, du brut stationnaire à l'Alpha net du scénario.	21
10.1. Distribution de l'Alpha net du scénario sur 250 000 tirages déclarés.	24
10.2. Sensibilité de l'Alpha net du scénario à une augmentation du multiplicateur de biais caché.	25
11.1. Alpha net du scénario dans l'écologie institutionnelle contradictoire déclarée.	26
12.1. Atlas exact des décisions neuronales-symboliques dans la trousse de simulation.	28
14.1. Synthèse stratégique et physique. Les quatre perspectives analytiques contraignent le même candidat sous le contrôle du client, tandis que l'autorité effective demeure A0.	34
14.2. Conception du mécanisme du jeu de preuve. La séparation des rôles, les preuves nouvelles, les clés distinctes, le préengagement et une Chronique à ajout seulement augmentent le coût attendu de la manipulation et réduisent le gain de collusion.	35
14.3. Paysage décisionnel d'énergie libre de Gibbs. La qualification locale peut constituer un minimum métastable; une barrière de vérification sur preuves nouvelles la sépare de la preuve indépendante. Les règles absolues de preuve et d'autorité demeurent discrètes.	37
14.4. Formation hamiltonienne et discontinuités institutionnelles. La recherche réversible de candidats a lieu avant le figement exact. La preuve indépendante évalue l'objet figé sur des preuves nouvelles sans effectuer de mise à jour des paramètres.	38
14.5. Diagramme de phases institutionnelles. Selon les hypothèses de simulation, le candidat actuel se trouve dans la région prête pour la preuve. Une hausse de l'incertitude peut ramener l'institution à la formation ou imposer l'arrêt. L'autorité demeure une coordonnée orthogonale.	39
15.1. Grande synthèse. Quatre lentilles analytiques agissent sur le même candidat figé exact tandis que la frontière des allégations et l'autorité A0 demeurent invariantes.	43
15.2. Profil informationnel et thermodynamique de la preuve. La qualification locale réduit l'incertitude, mais la contraction probante la plus importante survient uniquement après le franchissement de la barrière des preuves nouvelles.	44

15.3. Carte de phases contrainte. L'énergie libre classe les états admissibles ; les contrôles durs
de preuve et d'autorité déterminent les états qui appartiennent à l'ensemble faisable. . . 45

15.4. Écologie évolutive illustrative des vérificateurs. Des incitatifs appropriés et une intensité
d'audit suffisante élargissent le bassin d'attraction de la vérification véridique. 46

15.5. Accessibilité sûre. Le candidat exact peut passer de la qualification locale vers la preuve
indépendante sans passer à une coordonnée d'autorité supérieure. 47

16.1. Programme proposé de preuve fantôme indépendante A0 de 90 jours. 50

Liste des tableaux

1.	Droits de décision et état probant actuel.	i
2.	Parcours de lecture conçus selon le rôle.	vi
2.1.	Allégations autorisées par les preuves exactes.	6
4.1.	Liaisons principales des artefacts.	9
5.1.	Enveloppe de portefeuille constituée.	11
7.1.	Tournoi de candidats hérité de l'étude locale exacte.	15
9.1.	Déductions à dénominateur complet.	21
10.1.	Résumé Monte-Carlo.	24
11.1.	Écologie de stress et réponse institutionnelle prescrite.	27
12.1.	Décisions bornées représentatives.	29
13.1.	Contrôles minimaux du conseil avant toute décision d'admission.	32
14.1.	Correspondance entre physique stratégique, théorie des jeux et gouvernance institutionnelle.	40
15.1.	Signification décisionnelle des grandeurs stratégiques et physiques.	45
15.2.	Logique de refus destinée au conseil.	48
17.1.	Matrice de décision pour la prochaine étape institutionnelle.	53
B.1.	Distributions d'incertitude divulguées.	55
C.1.	Résumé des sondes du modèle exact.	56
D.1.	Résumé du contrat de preuve.	57
E.1.	Principaux artefacts publiés.	58
F.1.	Liste de vérification fondée sur les artefacts.	59

Fondements de l'intelligence de mission assujettie à la preuve

Constitution de mission, sémantique des allégations, identité des artefacts et candidat exact sous le contrôle du client.



LA VALEUR PEUT JUSTIFIER LA PREUVE; ELLE N'ACHÈTE JAMAIS LA PREUVE, L'ADMISSION OU L'AUTORITÉ.

1 Thèse institutionnelle

1.1 L'unité de valeur n'est pas le modèle

Les organisations modernes peuvent louer des modèles toujours plus capables, mais une capacité louée ne constitue pas encore un actif institutionnel. Une institution exige une mission durable, des comparateurs constitués, une garde des preuves, des droits de décision explicites, une séparation des rôles, un mécanisme de retour arrière et un dossier retraçant la manière dont les allégations ont été acquises. La proposition centrale du présent rapport n'est donc pas qu'un petit réseau neuronal soit « superintelligent ». Elle est qu'un composite sous le contrôle du client peut devenir institutionnellement important lorsqu'il transporte, dans un même objet vérifiable, une politique apprise, des contrôles symboliques, des preuves, des obligations de preuve, des limites d'autorité et une mémoire prête pour la succession.

L'objet mis à l'essai est volontairement étroit. Il vise l'intervention sur les paiements fournisseurs dans un portefeuille fictif. Sa politique apprise interne associe six observations normalisées à trois actions bornées. Sa couche symbolique peut encore restreindre la proposition, et l'enveloppe opérationnelle empêche toute action externe. Cette étroitesse constitue un avantage : elle rend l'allégation falsifiable, le comparateur constitutif, l'objectif mesurable et l'autorité séparable de la capacité.

Principe 1.1 (La capacité peut se composer ; l'autorité ne doit pas s'hériter automatiquement). Une amélioration de la qualité prédictive ou décisionnelle ne justifie pas, à elle seule, un statut de preuve indépendante, une admission en production ou une autorité conséquente. Chacune de ces coordonnées exige des preuves distinctes et une décision responsable.

1.2 Thèse de décision

Le candidat local exact est suffisamment prometteur pour financer un programme de preuve indépendante. Trois caractéristiques rendent cette décision rationnelle :

1. l'écart local est mesuré par rapport au comparateur le plus fort, et non seulement par rapport au processus en place ;
2. la thèse de portefeuille demeure positive après une perte de transfert explicite et 103 M \$ CA de déductions à dénominateur complet ;
3. le système consigne les écologies négatives et s'arrête dans les conditions extrêmes au lieu de transformer chaque état en récit positif.

La décision n'est pas de déployer. Elle consiste à acheter de l'information : une preuve fantôme A0 de 90 jours sous gouvernance indépendante, destinée à déterminer si le candidat exact figé conserve une supériorité propre à la mission sur des preuves nouvelles.

Action institutionnelle recommandée

Figurer les octets et les poids exacts du candidat; nommer un gardien indépendant des preuves et un vérificateur; authentifier le processus en place et les comparateurs Bêta; préengager la métrique principale, les déductions et le seuil; exécuter une évaluation fantôme A0 de 90 jours; conserver chaque échec et chaque abstention; et maintenir les décisions d'admission en production et d'autorité à l'extérieur de l'équipe de vérification.

1.3 Questions de recherche et résultats falsifiables

L'étude met à l'épreuve quatre propositions.

1. **Proposition relative aux artefacts.** Le candidat, l'état de référence, les preuves, la politique et le statut de preuve peuvent être liés dans un seul objet portable dont l'identité survit à la réouverture des octets exacts et aux contrôles d'altération.
2. **Proposition de valeur.** Le signal local peut soutenir une thèse positive de souscription à l'échelle du portefeuille après des réductions de transfert et des déductions à dénominateur complet.
3. **Proposition de robustesse.** La thèse de valeur n'est pas uniquement un artefact d'un environnement stationnaire; elle demeure pertinente pour la décision dans les régimes déclarés d'incertitude et de stress, tout en conservant les échecs visibles.
4. **Proposition de gouvernance.** Une politique apprise utile peut fonctionner sous un bouclier symbolique déterministe et une autorité effective A0 sans acquérir implicitement le droit d'agir.

Un résultat négatif pour l'une de ces propositions réduirait la recommandation. Une défaillance de l'identité de l'artefact bloquerait tout travail ultérieur. Un scénario prudent de valeur négative rendrait économiquement injustifié le programme de preuve. Un effondrement sous une faible perte de transfert signifierait une fragilité. Une défaillance de la limite d'autorité invaliderait l'architecture institutionnelle.

2 Système d'allégations et définitions opérationnelles

2.1 Échelle des preuves

Ce rapport utilise une échelle d'allégations plutôt qu'une seule étiquette de préparation. Soit l'état du système

$$S = (C, P, D, A), \quad (2.1)$$

où C désigne la formation de la capacité, P le statut de preuve, D le statut de déploiement ou d'admission et A l'autorité effective. Ces coordonnées ne s'impliquent volontairement pas :

$$C_{\text{qualification locale}} \not\Rightarrow P_{\text{preuve indépendante}}, \quad P_{\text{preuve indépendante}} \not\Rightarrow D_{\text{admis}}, \quad D_{\text{admis}} \not\Rightarrow A_{>0}. \quad (2.2)$$

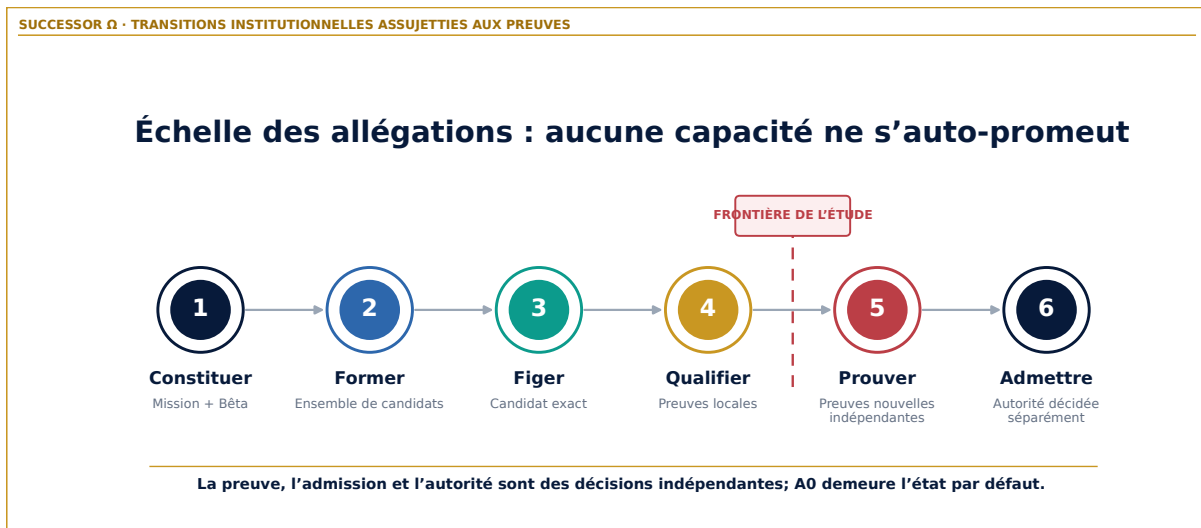


FIGURE 2.1. – Échelle des allégations. La qualification locale ne constitue ni preuve indépendante, ni admission, ni autorité.

2.2 Définitions

Définition 2.1 (Processus en place). Le processus en place est le processus organisationnel actuellement constitué ou le dossier d'actions importé par rapport auquel la valeur de la mission est mesurée.

Définition 2.2 (Bêta). Bêta est la meilleure capacité externe crédible et remplaçable constituée avant la formation du candidat. Dans l'étude locale, Bêta est un instantané figé d'actions. Les actions four-

nies sont utiles comme dossier de comparaison, mais leur provenance externe n'est pas authentifiée indépendamment.

Definition 2.3 (Plancher de référence). Étant donné les valeurs nettes V_I et V_β du processus en place et de Bêta figé,

$$B^* = \max\{V_I, V_\beta\}. \quad (2.3)$$

Un candidat doit dépasser le comparateur le plus fort, et non seulement le processus en place.

Definition 2.4 (Alpha* du scénario local). Pour une valeur V_Ω du candidat figé,

$$\alpha_{\text{local}}^* = V_\Omega - B^*. \quad (2.4)$$

L'astérisque indique une estimation bornée et propre au protocole. Il ne s'agit pas d'une valeur d'affaires réalisée.

Definition 2.5 (Candidat spécialiste). Un candidat spécialiste est le composite propre à la mission qui réunit modèle, programmes symboliques, liaisons de preuves, état de référence, essais, statut de preuve, identité, enveloppe d'autorité et Chronique. Le réseau neuronal seul n'est pas le candidat.

Definition 2.6 (Candidat de preuve d'ASI spécialiste). Un candidat de preuve d'ASI spécialiste est un candidat spécialiste figé dont les preuves locales et la justification de valeur institutionnelle justifient une preuve indépendante. Le terme n'affirme pas qu'une supériorité indépendante a déjà été établie.

Definition 2.7 (ASI spécialiste). Le terme ASI spécialiste est réservé ici à un candidat exact figé dont la supériorité propre à la mission par rapport aux comparateurs constitués a été établie au moyen de preuves nouvelles sous garde indépendante, de métriques préengagées, d'une falsification et d'un reçu valide du vérificateur. Le présent rapport n'accorde pas cette désignation.

Definition 2.8 (SUCCESSOR Ω). Un SUCCESSOR Ω est une continuation admise institutionnellement d'une intelligence de mission démontrée, avec identité, mémoire, preuves, politique, surveillance, retour arrière et contrôles de succession durables. La preuve, l'admission et l'autorité doivent être renouvelées ou héritées uniquement selon leurs propres règles.

2.3 Matrice des allégations autorisées

TABLE 2.1. – Allégations autorisées par les preuves exactes.

Allégation	Statut	Fondement
Candidat utile entraîné pour la mission	RÉUSSI	Objet entraîné exact, jeu de contrôle local, rejeu du modèle
Alpha* positif du scénario local	RÉUSSI	89 302 \$ CA au-dessus de Bêta figé
Valeur de scénario à l'échelle du portefeuille	RÉUSSI	Modèle fictif transparent de souscription
Alpha réel ou économies réalisées	NON ÉTABLI	Aucune mesure opérationnelle
Preuve indépendante	NON ÉTABLI	Même filiation de simulation ; aucune garde indépendante
Admission en production	NON ÉTABLI	Aucun environnement client admis
Autorité supérieure à A0	NON ÉTABLI	Enveloppe A0 et zéro effet externe
ASI spécialiste	NON ÉTABLI	Preuve indépendante non terminée
ASI générale	NON ÉTABLI	Hors de la mission et du périmètre des preuves

3 Travaux connexes et filiation conceptuelle

3.1 Systèmes neuronaux-symboliques et assurance à l'exécution

La recherche neuronale-symbolique combine des représentations apprises à des structures explicites de connaissance et de raisonnement (BESOLD et al., 2017). La présente conception s'inscrit dans cette tradition, mais confie à la couche symbolique des responsabilités institutionnelles : observations typées, bouclier de politique, calcul de preuve et refus d'autorité. L'apprentissage par renforcement sécuritaire au moyen de boucliers montre comment un mécanisme distinct peut bloquer des actions dangereuses de l'apprenant (ALSHIEKH et al., 2018); les architectures Neural Simplex séparent de façon analogue un contrôleur avancé d'un contrôleur de sécurité digne de confiance (PHAN et al., 2020). Les propositions d'IA à sûreté garantie insistent sur les modèles du monde, les spécifications de sûreté et la vérification (DALRYMPLE et al., 2024). Cette étude ne revendique aucune garantie formelle de sûreté; elle démontre une architecture concrète où des contraintes symboliques peuvent déclasser les propositions apprises tout en maintenant l'autorité effective à A0.

3.2 Évaluation comparative, intégrité du jeu de contrôle et changement de domaine

Le candidat local ressemble davantage à une politique de bandit contextuel qu'à un contrôleur autonome à long horizon (LI et al., 2010; SUTTON et BARTO, 2018). Le schéma reset/step inspiré de Gym fournit une interface standardisée pour une interaction répétable (BROCKMAN et al., 2016; TOWERS et al., 2024). La réutilisation adaptative d'un jeu de contrôle et les fuites peuvent invalider les allégations statistiques (DWORK et al., 2014; DWORK et al., 2015; KAPOOR et NARAYANAN, 2023); le protocole local sépare donc entraînement, sélection en développement et évaluation scellée après figement.

Le transfert à l'échelle du portefeuille n'est pas présumé invariant. Les travaux sur la généralisation hors domaine montrent que la robustesse dans des environnements non observés ne peut être déduite des seules performances internes (ARJOVSKY et al., 2019; GULRAJANI et LOPEZ-PAZ, 2021). La simulation traite donc le transfert comme un facteur explicite de rétention, effectue une analyse de sensibilité au biais caché et soumet la justification de valeur à des écologies de dérive et d'adversité.

3.3 Risque, incertitude et preuves de gouvernance

Le modèle de portefeuille présente des quantiles, la masse de probabilité au-dessus d'un seuil et une valeur conditionnelle dans la queue inférieure plutôt qu'un unique point estimé, conformément à une analyse

sensible au risque telle que la valeur à risque conditionnelle (ROCKAFELLAR et URYASEV, 2000). L'analyse de sensibilité rend visibles les hypothèses structurelles au lieu de les enfouir dans un seul score (SALTELLI et al., 2008). La preuve indépendante proposée pourra également recourir à des inférences uniformes dans le temps lorsque les preuves arriveront séquentiellement (HOWARD et al., 2021).

L'architecture de rapport s'inspire des fiches de modèle, des fiches de données, des cadres d'audit interne, des grilles de préparation ML et des recommandations REFORMS (MITCHELL et al., 2019; GEBRU et al., 2021; RAJI et al., 2020; BRECK et al., 2017; KAPOOR, CANTRELL et al., 2024). La séparation entre capacité, preuve, admission et autorité est également cohérente avec les principes de gestion du risque qui exigent contexte, mesure, gouvernance et gestion continue plutôt qu'un seul score de modèle (NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 2023). La dette technique cachée demeure pertinente : un modèle performant intégré à un système opaque ou non vérifiable n'est pas un actif institutionnel (SCULLEY et al., 2015).

3.4 Objets porteurs de preuves et historique résistant à l'altération

Le code porteur de preuve a établi le schéma selon lequel un producteur fournit un certificat vérifiable par machine que le consommateur valide par rapport à une politique (NECULA, 1997). L'objet entraîné de cette étude n'est pas du code porteur de preuve au sens formel, mais il est porteur de preuves : il lie mission, modèle, programmes, essais, statut et autorité à des empreintes. Les dossiers liés par hachage s'inscrivent dans les idées classiques d'horodatage résistant à l'altération (HABER et STORNETTA, 1991). SHA-256 (NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 2015) et les conventions de JSON canonique (RUNDGREN et al., 2020) fournissent des identités reproductibles, tout en reconnaissant explicitement qu'une chaîne locale non ancrée peut être copiée ou bifurquée.

4 Filiation exacte des artéfacts et reproductibilité

4.1 Intrants liés

La simulation scientifique commence par quatre intrants R2 exacts. Le successeur opérationnel actuellement commercialisable est la version R3, liée par une empreinte distincte ; elle corrige le routage de la Console sans modifier rétroactivement les preuves de recherche figées.

TABLE 4.1. – Liaisons principales des artéfacts.

Artéfact	SHA-256
Trousse acheteur R2 liée à la recherche	f7747d176def4594b1f65f642a54c8ba49d8396747f736eaced1ab700dd27e4e
Trousse acheteur R3 opérationnelle actuelle	3f15b848306e96d2825fd2a5ec68b112dd13b49806685e93986259e2498ac67a
Candidat spécialiste entraîné	142706131d9b017e0c20c5f09a127cf66d700f0a7989bd9f4f302d29afddb915
Demande de preuve indépendante	c30cbf2440d9b7648ba8622b31fedd8e70087ac3e587f1140fcfd8d2d9f277be
Preuves de l'API A0	b5c7bd14c35317f9d12a223cbdd69ae08400e2b9b03a69c40ca7ddf462526b66

La trousse acheteur réussit la validation CRC du ZIP avec 270 entrées, aucune entrée dupliquée et aucun chemin dangereux. Le modèle candidat possède une matrice de poids entrée-couche cachée de 8×6 et une matrice couche cachée-sortie de 3×8 . Le rejeu du candidat exact est déterministe.

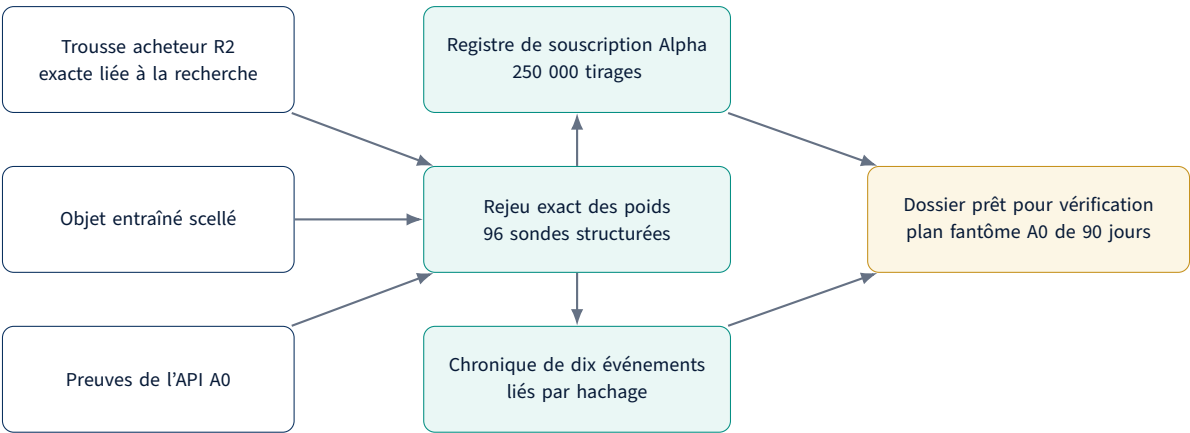


FIGURE 4.1. – Filiation des artéfacts, des intrants exacts au dossier de preuve indépendante.

4.2 Contrôles de reproductibilité

La trousse de version comprend des résumés lisibles par machine, le code source, les dossiers de sondes du modèle, les tirages Monte-Carlo, l'écologie de stress, le registre Alpha, la Chronique, la constitution de mission, le plan de preuve et les empreintes complètes. La trousse source du présent article comprend les données exactes utilisées pour chaque figure et tableau.

Contrat de reproductibilité

Un évaluateur doit pouvoir : (1) vérifier chaque empreinte d'intrant ; (2) régénérer les figures ; (3) reconstruire les trois éditions de l'article ; (4) confirmer la structure et le rendu des PDF ; (5) reproduire le résumé Monte-Carlo à partir du NPZ fourni ; et (6) détecter toute modification au moyen de l'inventaire SHA-256 publié.

4.3 Sémantique de la Chronique

La Chronique de simulation contient dix événements à ajout seulement. Soit r_t le dossier canonique à la séquence t et h_{t-1} l'empreinte précédente. La chaîne est

$$h_t = \text{SHA256}(\text{JCS}(r_t \parallel h_{t-1})), \quad h_0 = 0^{64}. \quad (4.1)$$

La tête finale est

0f22eef59acdd5b1c544ba4a337719d046964e52670e053c92ea0ab48ee3bf19.

La chaîne établit la cohérence de la séquence locale présentée, et non son unicité mondiale, une heure digne de confiance ou un ancrage externe.

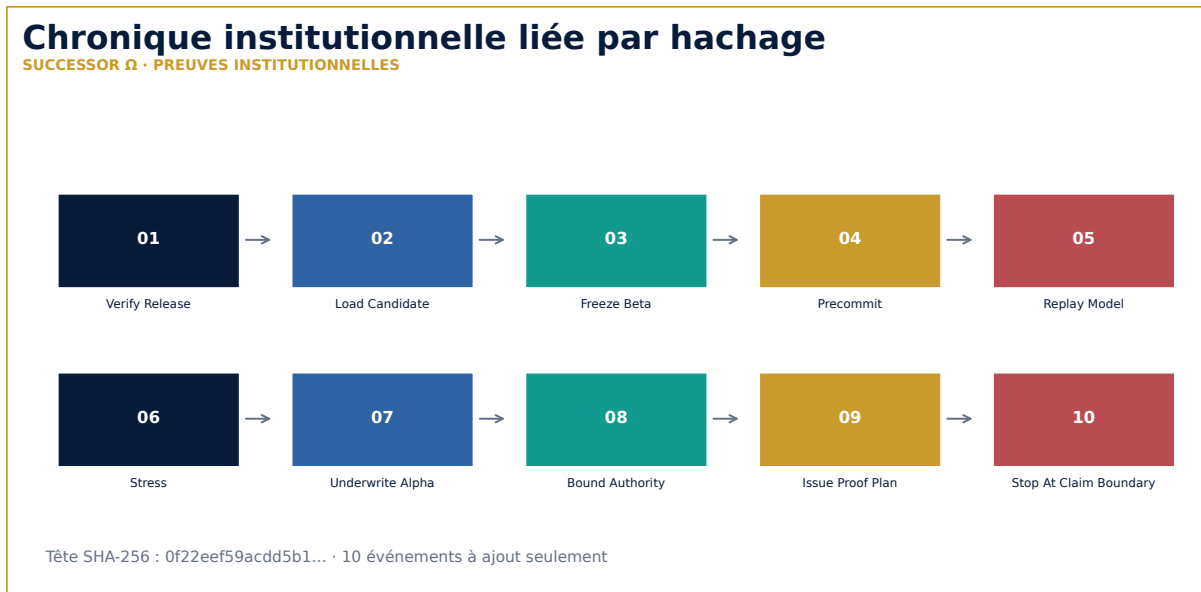


FIGURE 4.2. – Chronique de simulation à dix événements.

5 Constitution de mission et enveloppe de portefeuille

5.1 Institution fictive et mission

L’institution fictive est le Groupe Industries Stratégiques Aurelius. L’utilisatrice technique simulée est la Dre Amara Laurent, scientifique en chef des systèmes d’intelligence institutionnelle. La mission consiste à protéger un portefeuille fictif annuel de paiements fournisseurs de 48,8 G \$ CA au moyen de recommandations réversibles et liées à des preuves qui améliorent la valeur nette de prévention et de continuité par rapport au meilleur du processus en place et d’un comparateur Bêta figé.

TABLE 5.1. – Enveloppe de portefeuille constituée.

Dimension	Valeur constituée
Flux annuel de paiements fournisseurs	48,8 G \$ CA
Occasions annuelles pondérées de décision	1 000 000
Fournisseurs	72 800
Entités opérationnelles	41
Territoires	26
Autorité effective	A0 : observer, inspecter et proposer seulement
Effets externes	0

5.2 Contraintes non négociables

1. Aucun paiement, blocage, suspension de fournisseur, changement de compte ou modification du grand livre ne peut être exécuté.
2. L’approbation humaine demeure obligatoire pour toute décision conséquente.
3. Les octets et les poids du candidat, les instantanés des comparateurs, les métriques, les seuils et les déductions sont figés avant la preuve sur données nouvelles.
4. Tout entraînement, ajustement de seuil ou changement de règle sur le jeu de contrôle de preuve est interdit.
5. Le coût complet, la réserve de risque, la charge humaine, la réserve d’imitation et la réserve de dérive demeurent dans le dénominateur.
6. Chaque échec, abstention, écologie négative et événement défavorable est conservé.

7. La provenance des comparateurs et les justificatifs du vérificateur doivent être authentifiés avant toute allégation de preuve indépendante.

Limite d'autorité

La constitution de mission peut décrire un plafond éventuel d'action, mais l'état réel demeure A0. Les propositions internes telles qu'« expérience réversible » sont des sorties informationnelles ; le moteur fourni ne possède aucun droit de production et ne produit aucun effet externe.

5.3 Pourquoi l'intégrité du capital fournisseurs est une mission exigeante

L'intervention sur les paiements fournisseurs combine coûts asymétriques, incertitude d'identité, modes de défaillance corrélés, continuité opérationnelle et contraintes d'examen humain. Un faux positif peut retarder un paiement légitime, nuire à la relation avec un fournisseur ou interrompre la production ; un faux négatif peut laisser passer un paiement en double, anormal ou compromis. La meilleure politique doit donc optimiser un objectif de mission complet, et non un simple score de détection.

La mission est également contradictoire. Les fournisseurs et les attaquants peuvent s'adapter ; les données maîtres peuvent être compromises ; les preuves peuvent arriver tardivement ; la capacité des réviseurs peut diminuer ; et un service Bêta externe peut progresser. Ces conditions motivent l'écologie de stress du chapitre 11 et le modèle explicite de transfert du chapitre 10.

6 Architecture institutionnelle neuronale-symbolique

6.1 Boucle de fonctionnement

L'architecture combine une politique apprise, des programmes symboliques typés et une gouvernance assujettie à la preuve. Le parcours est le suivant :

Constituer → Observer → Entraîner → Falsifier → Figurer → Évaluer → Souscrire → Prouver → Admettre → Borner l'autorité.

La composante apprise propose une action à partir de six observations bornées. La composante symbolique interprète cette proposition selon des règles de preuve, de réversibilité, de parties prenantes, de changement d'identité et d'autorité. La composante de preuve calcule la valeur relative aux comparateurs et le statut. La composante d'autorité refuse toute voie supérieure à A0 dans la version mise à l'essai.

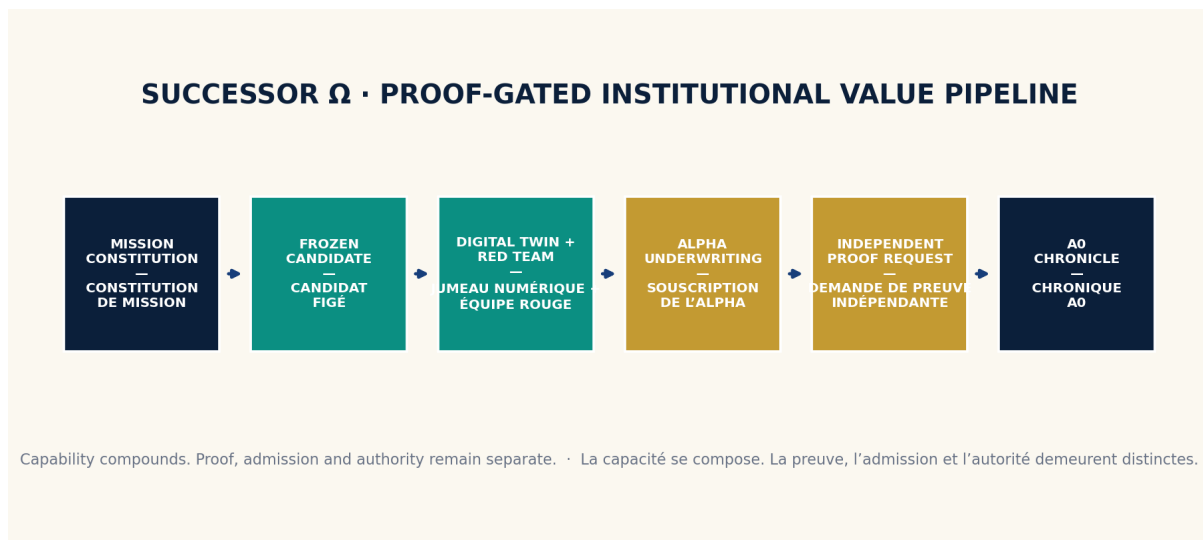


FIGURE 6.1. – Pipeline de valeur institutionnelle représenté dans la trousse exacte de simulation.

6.2 Le spécialiste est le composite

Un modèle ne constitue qu'un élément de l'objet. Le candidat spécialiste scellé transporte également :

- la constitution de mission et le registre de référence ;
- les liaisons de preuves et l'historique des partitions ;

- le tournoi de candidats et les poids sélectionnés ;
- les programmes Monde, Politique, Preuve et Autorité ;
- les essais locaux, l'historique de falsification et le registre Alpha ;
- un pare-feu A0 et un adaptateur d'actions bornées ;
- une Chronique, une fiche signée et un coffre chiffré de clé du détenteur ;
- une demande de preuve indépendante.

Cette composition transforme un prédicteur remplaçable en artéfact institutionnel portable. L'objet n'établit toujours pas le titre juridique, le contrôle exclusif, une heure digne de confiance, l'identité externe ou une preuve indépendante ; ces allégations exigent des institutions supplémentaires.

6.3 Capacité, preuve, admission et autorité

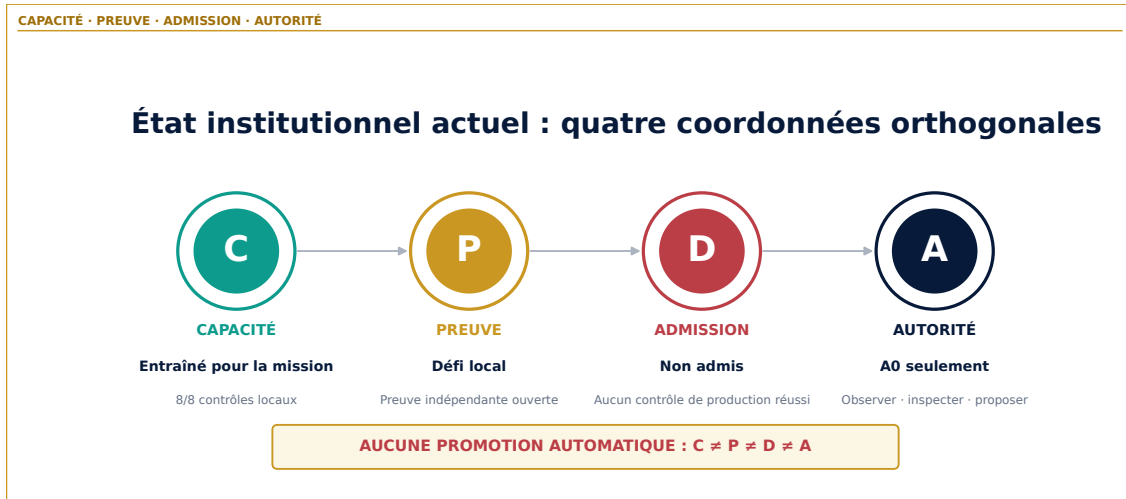


FIGURE 6.2. – Séparation de la capacité, de la preuve, de l'admission et de l'autorité dans la simulation.

L'état actuel peut être résumé comme suit :

$$S_{\text{actuel}} = (C_{\text{qualifié localement}}, P_{\text{ouverte}}, D_{\text{non admis}}, A_0). \quad (6.1)$$

L'étape suivante proposée ne modifie que la coordonnée de preuve : une preuve indépendante réussie produirait $P_{\text{preuve indépendante}}$, tandis que D et A demeureraient inchangées jusqu'à des décisions distinctes.

7 Conception expérimentale et candidat exact

7.1 Filiation des preuves locales

L'étude locale utilise 1 000 cas synthétiques de paiements fournisseurs divisés en 600 cas d'entraînement, 200 cas de développement et 200 cas de contrôle scellés. Les actions du processus en place et de Bêta figé sont présentes pour chaque ligne. Le jeu de contrôle demeure à l'extérieur de l'entraînement et de la sélection en développement.

Trois candidats neuronaux-symboliques matériellement distincts reçoivent au total 27 000 mises à jour de paramètres. Le candidat figé sélectionné est une politique compacte $6 \rightarrow 8 \rightarrow 3$ comportant 83 paramètres appris. Le candidat n'est évalué qu'après son figement.

TABLE 7.1. – Tournoi de candidats hérité de l'étude locale exacte.

Candidat	Mises à jour	Topologie	Paramètres	Actions exactes en développement
Explorateur compact	8 000	$6 \rightarrow 8 \rightarrow 3$	83	92%
Synthétiseur équilibré	10 000	$6 \rightarrow 12 \rightarrow 3$	123	90%
Gardien conservateur	9 000	$6 \rightarrow 10 \rightarrow 3$	103	66%

7.2 Rejeu du modèle exact

La simulation de pointe recharge l'objet sous le contrôle du client, valide les empreintes de l'objet et du modèle, extrait les poids exacts et les rejoue sur 96 sondes structurées. La grille de sondes fait varier les signaux d'anomalie, de doublon, de changement de compte, de force des preuves, de réversibilité et de risque pour les parties prenantes. La sortie neuronale est interprétée par un bouclier symbolique déterministe.

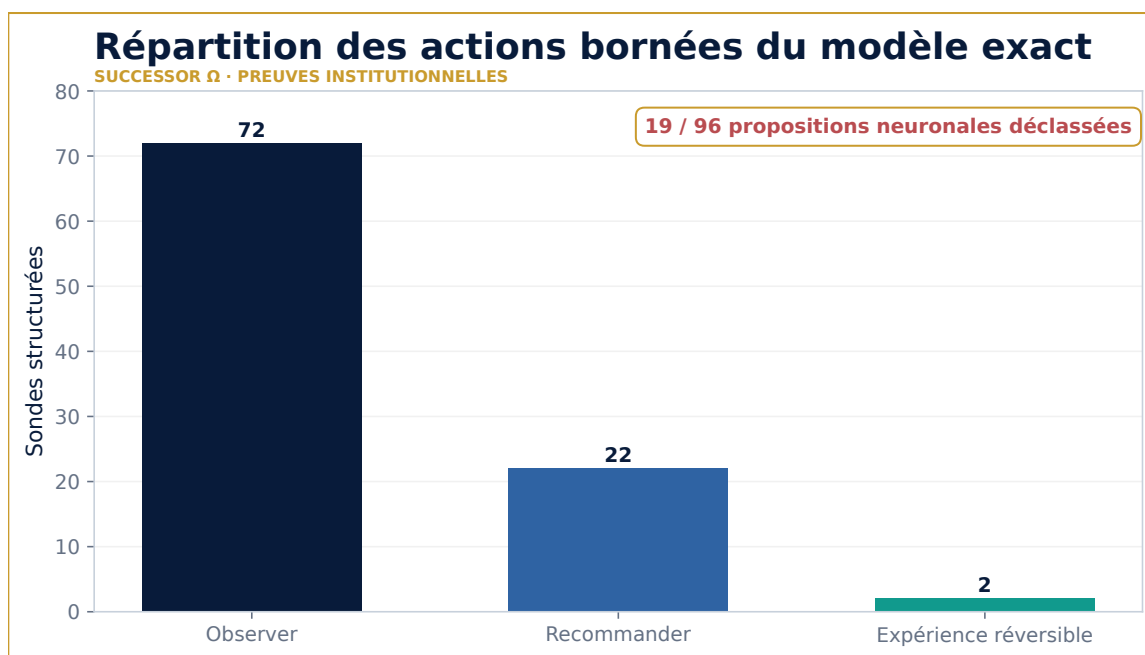


FIGURE 7.1. – Répartition des actions bornées lors du rejou du modèle exact.

Les sorties bornées comprennent 72 observations, 22 recommandations et deux propositions d'expérience réversible. Dix-neuf propositions neuronales sont déclassées par au moins une règle symbolique. Le plancher de preuve est fréquemment actif, ce qui montre que l'objet ne se contente pas de maximiser l'intensité d'intervention.

7.3 Contrôles absolus locaux

La qualification locale exige la réussite des huit contrôles absolus. Ceux-ci couvrent notamment la couverture des comparateurs, la séparation du jeu de contrôle, le figement exact du candidat, l'absence de violation de sécurité déclarée, un écart positif par rapport au comparateur, le soutien par une borne inférieure, l'arithmétique complète et la préservation de l'autorité A0. La réussite de tous les contrôles autorise l'expression « candidat qualifié localement », et non « preuve indépendante ».

8 Qualification locale et formation du candidat

8.1 Résultat local exact

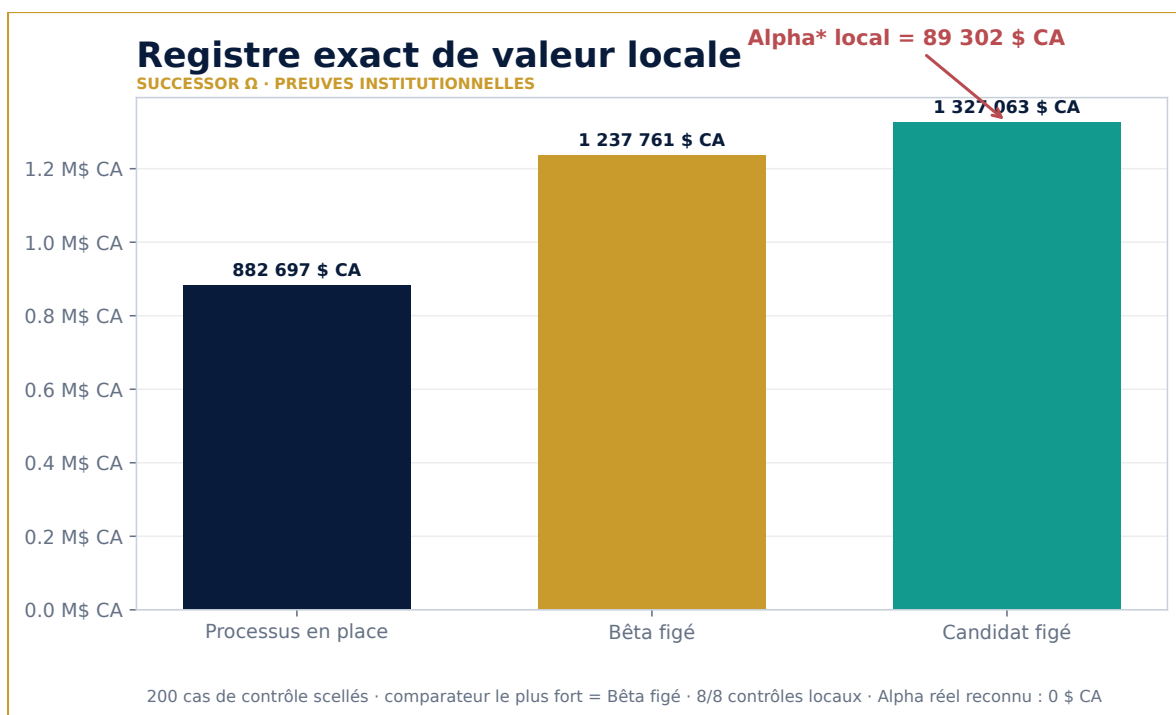


FIGURE 8.1. – Registre exact de valeur locale sur le jeu de contrôle scellé de 200 cas.

Le processus en place produit 882 697 \$ CA, Bêta figé produit 1 237 761 \$ CA et le candidat figé produit 1 327 063 \$ CA. Bêta est le comparateur le plus fort, donc

$$\alpha_{\text{local}}^* = 1\,327\,063 - 1\,237\,761 = 89\,302 \text{ \$ CA.} \quad (8.1)$$

L'erreur-type appariée est de 20 967 \$ CA. La borne inférieure normale de sensibilité est de 48 208 \$ CA, et la borne inférieure du bootstrap déterministe atteint 49 340 \$ CA. Le candidat obtient 91,5% d'actions exactes, zéro expérience nuisible déclarée et huit contrôles locaux réussis sur huit.

Ce que le résultat local permet d'affirmer

Un Alpha* positif, borné et reproductible du scénario local pour le candidat exact figé par rapport au comparateur fourni le plus fort. Il permet également de demander légitimement si le signal se transfère. Il ne permet pas de reconnaître une valeur réalisée, d'affirmer une preuve indépendante ou d'accorder une autorité.

8.2 Pourquoi le comparateur le plus fort compte

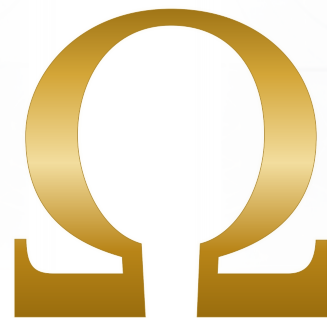
Une défaillance fréquente de l'évaluation de l'IA consiste à comparer un nouveau système uniquement à un processus en place inefficace en ignorant une solution externe crédible. Ici, Bêta dépasse le processus en place de 355 064 \$ CA ; le candidat doit donc dépasser Bêta pour créer de l'Alpha. Cette règle rend l'estimation pertinente pour une organisation qui pourrait louer ou acheter l'alternative externe.

8.3 Pourquoi le succès local demeure insuffisant

Les données sont synthétiques, la provenance du comparateur n'est pas authentifiée indépendamment, la filiation de simulation n'est pas indépendante sur le plan organisationnel, et l'environnement ne peut représenter tous les coûts opérationnels ou toutes les réponses stratégiques. Ces limites ne rendent pas le résultat local inutile. Elles déterminent l'obligation de preuve suivante.

Alpha, incertitude et robustesse contradictoire

Valeur à dénominateur complet, sensibilité au transfert, écologies de stress, assurance à l'exécution et frontière de refus.



LA VALEUR PEUT JUSTIFIER LA PREUVE; ELLE N'ACHÈTE JAMAIS LA PREUVE, L'ADMISSION OU L'AUTORITÉ.

9 Souscription du portefeuille à dénominateur complet

9.1 Modèle d'échelle

Le portefeuille fictif comprend un million d'occasions annuelles pondérées de décision. Par rapport au jeu de contrôle local de 200 cas, le facteur d'échelle est

$$s = \frac{1\,000\,000}{200} = 5\,000. \quad (9.1)$$

Une projection naïve sous stationnarité serait

$$G = \alpha_{\text{local}}^* s = 89\,302 \times 5\,000 = 446,510 \text{ M \$ CA.} \quad (9.2)$$

Cette valeur n'est pas retenue comme résultat décisionnel. Elle constitue un intrant contrefactuel brut de la souscription.

9.2 Rétention de transfert et dénominateur complet

Soit ρ la fraction de l'Alpha brut local retenue après transfert technique, couverture, adoption, adaptation contradictoire et qualité des preuves. Soient C, R, H, I, D le coût complet, la réserve de risque, la charge humaine, la réserve d'imitation et la réserve de dérive. L'Alpha net du scénario est

$$\alpha_{\text{portefeuille}} = \rho G - (C + R + H + I + D). \quad (9.3)$$

Le scénario central utilise $\rho = 0,72$ et les déductions du tableau 9.1.

TABLE 9.1. – Déductions à dénominateur complet.

Code	Interprétation	Déduction centrale
C	Coût complet de mise en œuvre, d'intégration, d'assurance et d'exploitation	32 M \$ CA
R	Réserve pour baisse, faux positifs, continuité et incidents	29 M \$ CA
H	Charge d'examen humain, d'enquête, d'escalade et de décision	14 M \$ CA
I	Réserve d'imitation et d'amélioration du comparateur	11 M \$ CA
D	Réserve de surveillance, de requalification et de dérive	17 M \$ CA
Total		103 M \$ CA

Le résultat central est

$$0,72 \times 446,510 \text{ M} - 103 \text{ M} = 218,4872 \text{ M } \$ \text{ CA.}$$

(9.4)

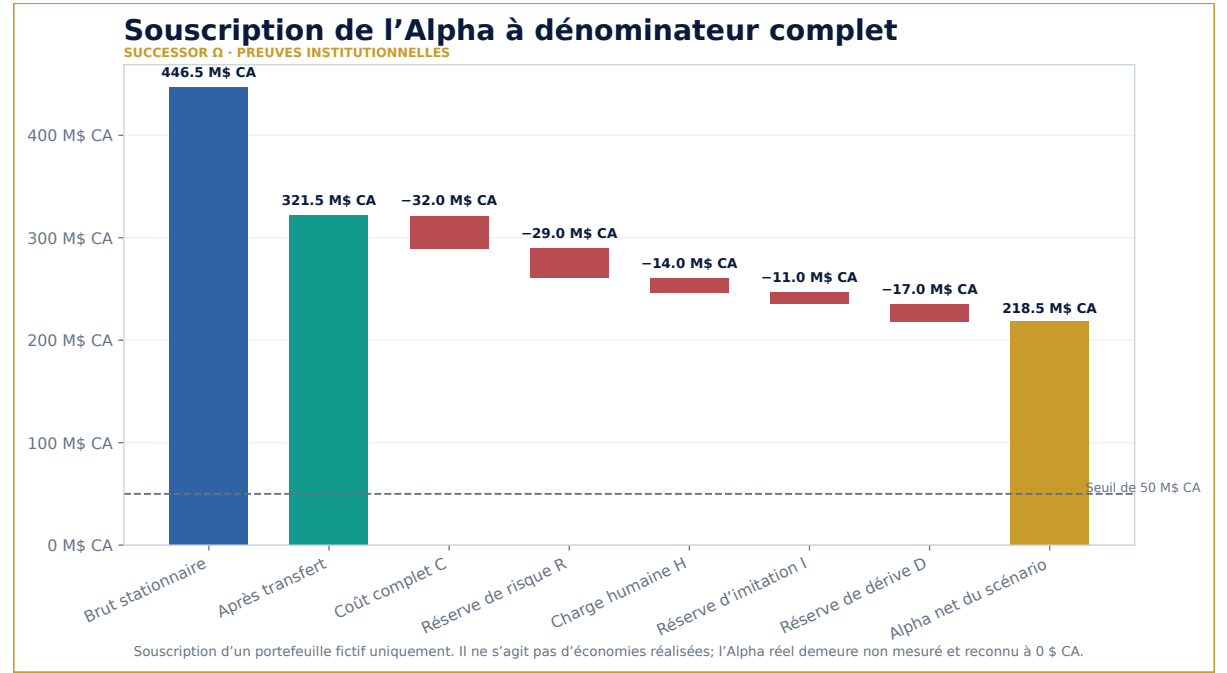


FIGURE 9.1. – Souscription à dénominateur complet, du brut stationnaire à l'Alpha net du scénario.

9.3 Scénario inférieur prudent

En utilisant la borne inférieure normale de sensibilité locale plutôt que l'estimation centrale, on obtient

$$\alpha_{\text{inférieur}} = 48\,208 \times 5\,000 \times 0,72 - 103 \text{ M} = 70,5488 \text{ M } \$ \text{ CA.}$$

(9.5)

Par rapport au seuil préengagé de 50 M \$ CA, la marge est de 20,5488 M \$ CA. Ce scénario inférieur constitue le résultat le plus solide pour le conseil parce qu'il commence par une borne d'incertitude locale et conserve le dénominateur complet.

La souscription n'est pas une réalisation

Le portefeuille est fictif. Le modèle d'échelle est transparent, mais repose sur des hypothèses. Aucun paiement opérationnel n'a été modifié, aucun recouvrement de trésorerie n'a été observé, et l'Alpha réel demeure non mesuré et reconnu à 0 \$ CA. La valeur du résultat est de justifier ou de rejeter l'investissement de preuve suivant.

9.4 Valeur économique de l'information

Le programme de preuve proposé possède de la valeur parce qu'il peut résoudre une incertitude à forte conséquence. Si la borne indépendante dépasse le seuil, l'institution obtient une preuve pour une décision d'admission. Si le signal s'effondre, elle évite un engagement de production beaucoup plus coûteux. Un programme discipliné de preuve n'est donc pas une dépense accessoire ; il constitue une option qui limite l'exposition du capital tout en préservant le potentiel d'un actif institutionnel précieux.

10 Incertitude distributionnelle et transfert causal

10.1 Modèle Monte-Carlo

La simulation exécute 250 000 tirages avec une graine fixe. La rétention brute est le produit de cinq facteurs bornés indépendants :

$$\rho = T K A E Q, \quad (10.1)$$

où T représente le transfert technique, K la couverture, A l'adoption, E la rétention face à l'adversité et Q la qualité des preuves. Les distributions déclarées sont

$$T \sim \text{Beta}(24, 4), \quad K \sim \text{Beta}(38, 2), \quad (10.2)$$

$$A \sim \text{Beta}(27, 3), \quad E \sim \text{Beta}(18, 4), \quad (10.3)$$

$$Q \sim \text{Beta}(29, 3). \quad (10.4)$$

Les cinq déductions utilisent des distributions triangulaires dont les valeurs minimale, modale et maximale sont présentées à l'annexe B.

Pour le tirage j ,

$$\alpha_j = G\rho_j - (C_j + R_j + H_j + I_j + D_j). \quad (10.5)$$

Il s'agit de distributions de scénario, et non de probabilités postérieures dérivées de résultats opérationnels observés.

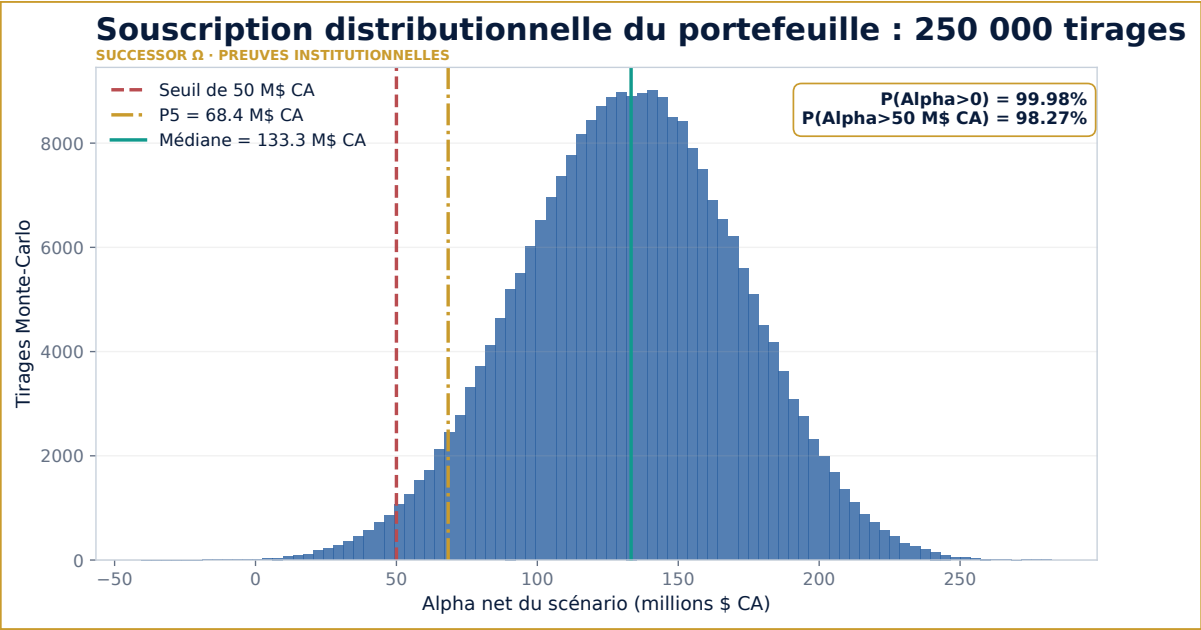


FIGURE 10.1. – Distribution de l’Alpha net du scénario sur 250 000 tirages déclarés.

10.2 Résultats distributionnels

TABLE 10.1. – Résumé Monte-Carlo.

Statistique	Valeur	Statistique	Valeur
Moyenne	132,843 M \$ CA	Médiane	133,267 M \$ CA
P1	42,148 M \$ CA	P5	68,357 M \$ CA
P10	82,429 M \$ CA	P25	106,423 M \$ CA
P75	159,616 M \$ CA	P90	182,736 M \$ CA
P95	196,066 M \$ CA	P99	220,147 M \$ CA
CVaR _{5%} de la queue inférieure	52,325 M \$ CA	Déductions moyennes	109,676 M \$ CA
$P(\alpha > 0)$	99,9792%	$P(\alpha > 50 \text{ M \$ CA})$	98,2728%

Le résultat P1 est inférieur au seuil, et la queue la plus basse comprend des tirages négatifs. Ce point est essentiel : le modèle ne garantit pas le succès. Il indique que, selon les distributions déclarées, la justification de valeur est suffisamment robuste pour mériter une preuve indépendante.

10.3 Sensibilité du transfert causal

Comme le transfert à l’échelle du portefeuille peut être confondu par des différences cachées, la simulation introduit un multiplicateur de biais caché γ et réduit la rétention selon

$$\rho(\gamma) = \frac{0,72}{\gamma^{0,72}}.$$

(10.6)

La justification de valeur demeure au-dessus du seuil de 50 M \$ CA jusqu'à $\gamma = 2,5$ et passe sous le seuil à $\gamma = 3,0$.

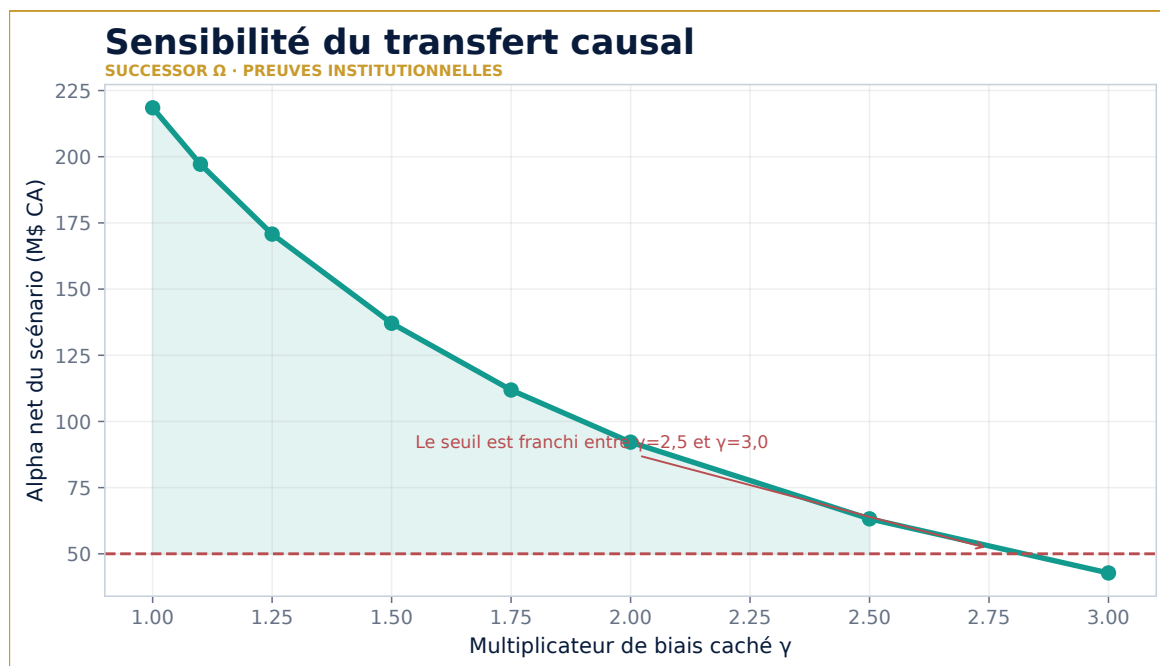


FIGURE 10.2. – Sensibilité de l'Alpha net du scénario à une augmentation du multiplicateur de biais caché.

Cette courbe ne constitue pas une preuve d'identification causale. Il s'agit d'une paramétrisation de stress transparente qui montre l'ampleur de la défaillance de transfert nécessaire pour invalider la thèse centrale d'investissement.

11 Écologie institutionnelle contradictoire

11.1 Conception

Une seule distribution peut masquer des modes de défaillance structurels. L'écologie de stress énumère donc treize états institutionnels nommés : stationnarité, souscription centrale, dérive modérée des covariables, dérive conceptuelle importante, adaptation stratégique des fournisseurs, amélioration du comparateur, réduction de la capacité des réviseurs, latence des preuves, choc du coût des faux positifs, attaque corrélée des données maîtres, conditions prudentes combinées, conditions extrêmes d'équipe rouge et arrêt de sûreté A0.

Chaque écologie modifie la rétention du transfert et, lorsqu'il y a lieu, ajoute une déduction précise. La réponse n'est pas choisie de manière à maximiser l'Alpha déclaré. Elle suit une règle de décision dépendante de l'état :

- poursuivre la preuve fantôme A0 lorsque l'Alpha net brut demeure au moins égal au seuil ;
- poursuivre un apprentissage restreint lorsque la valeur est positive mais inférieure au seuil ;
- arrêter et requalifier lorsque la valeur est non positive ou lorsque l'état de sûreté est déclenché.

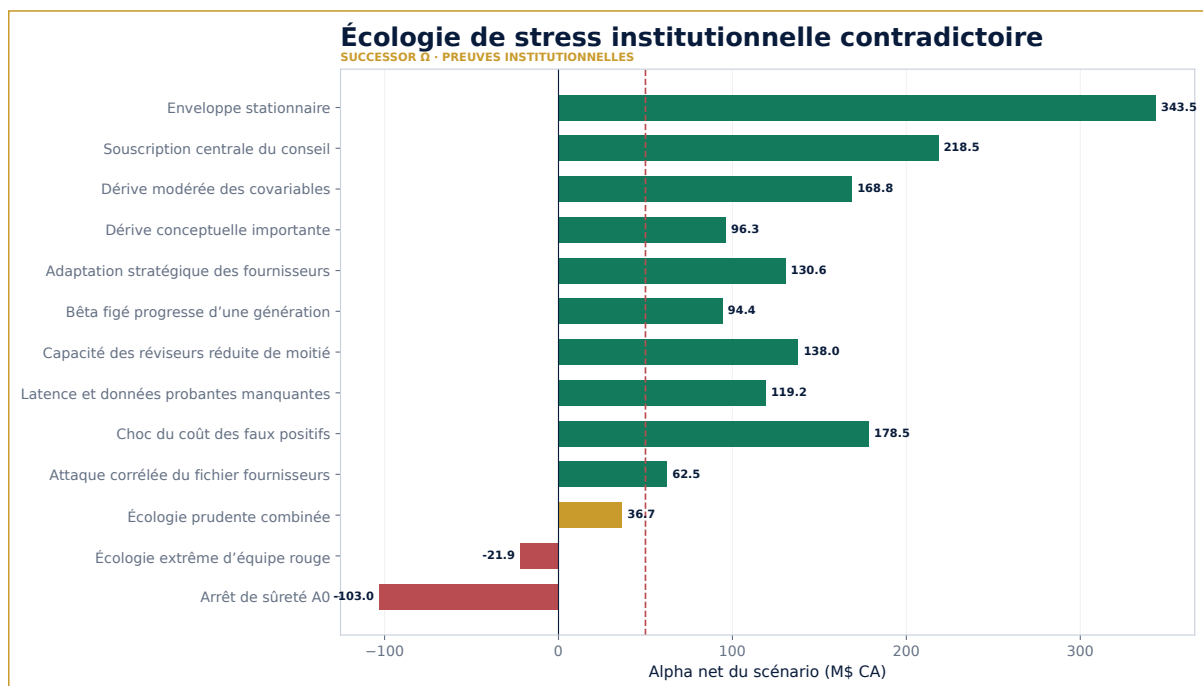


FIGURE 11.1. – Alpha net du scénario dans l'écologie institutionnelle contradictoire déclarée.

11.2 Interprétation

La plupart des écologies déclarées préservent une justification de valeur positive. La dérive conceptuelle importante, l'amélioration du comparateur et l'attaque corrélée du fichier fournisseurs réduisent sensiblement la marge. Les conditions prudentes combinées demeurent positives, mais passent sous le seuil du conseil, ce qui déclenche un apprentissage restreint plutôt qu'une proclamation de victoire. Les conditions extrêmes d'équipe rouge produisent un Alpha brut négatif; l'Alpha reconnu du scénario est alors nul, et la décision consiste à arrêter et à requalifier. L'état de sûreté A0 reconnaît lui aussi une valeur nulle.

Ne jamais faire disparaître l'échec par une moyenne

Les écologies négatives demeurent des preuves de premier ordre. Elles ne sont ni retirées comme valeurs aberrantes ni absorbées dans une moyenne positive. Toute version opérationnelle doit lier ses règles de surveillance et d'arrêt à ces catégories de défaillance avant toute admission.

11.3 Tableau complet des stress

TABLE 11.1. – Écologie de stress et réponse institutionnelle prescrite.				
ID	Écologie	Rétention	Alpha net brut	Décision
S0	Enveloppe stationnaire	1,00	343,510 M \$ CA	Poursuivre la preuve fantôme A0
S1	Souscription centrale du conseil	0,72	218,487 M \$ CA	Poursuivre la preuve fantôme A0
S2	Dérive modérée des covariables	0,62	168,836 M \$ CA	Poursuivre la preuve fantôme A0
S3	Dérive conceptuelle importante	0,48	96,325 M \$ CA	Poursuivre la preuve fantôme A0
S4	Adaptation stratégique des fournisseurs	0,55	130,581 M \$ CA	Poursuivre la preuve fantôme A0
S5	Bêta figé progresse d'une génération	0,46	94,395 M \$ CA	Poursuivre la preuve fantôme A0
S6	Capacité des réviseurs réduite de moitié	0,58	137,976 M \$ CA	Poursuivre la preuve fantôme A0
S7	Latence et données probantes manquantes	0,52	119,185 M \$ CA	Poursuivre la preuve fantôme A0
S8	Choc du coût des faux positifs	0,72	178,487 M \$ CA	Poursuivre la preuve fantôme A0
S9	Attaque corrélée du fichier fournisseurs	0,42	62,534 M \$ CA	Poursuivre la preuve fantôme A0
S10	Écologie prudente combinée	0,38	36,674 M \$ CA	Apprentissage restreint
S11	Écologie extrême d'équipe rouge	0,26	−21,907 M \$ CA	Arrêter et requalifier
S12	Arrêt de sûreté A0	0,00	−103,000 M \$ CA	Arrêter et requalifier

12 Moteur A0, bouclier symbolique et Chronique

12.1 Politique exacte dans des contraintes constitutionnelles

Le candidat n'est pas traité comme un agent libre. La politique neuronale propose l'une de trois actions bornées. Une couche symbolique déterministe vérifie la force des preuves, la réversibilité, le risque pour les parties prenantes, le changement d'identité du compte, la dérive et l'enveloppe d'autorité. Le résultat effectif demeure informationnel sous A0.

Sur 96 sondes structurées, 19 propositions neuronales sont modifiées par les règles symboliques. Une catégorie particulièrement importante apparaît lorsque la politique apprise propose une expérience réversible alors que les preuves ou les conditions relatives aux parties prenantes sont insuffisantes. La couche symbolique ramène le résultat à une recommandation ou à l'observation. Il s'agit d'une démonstration directe d'une capacité subordonnée à la politique.

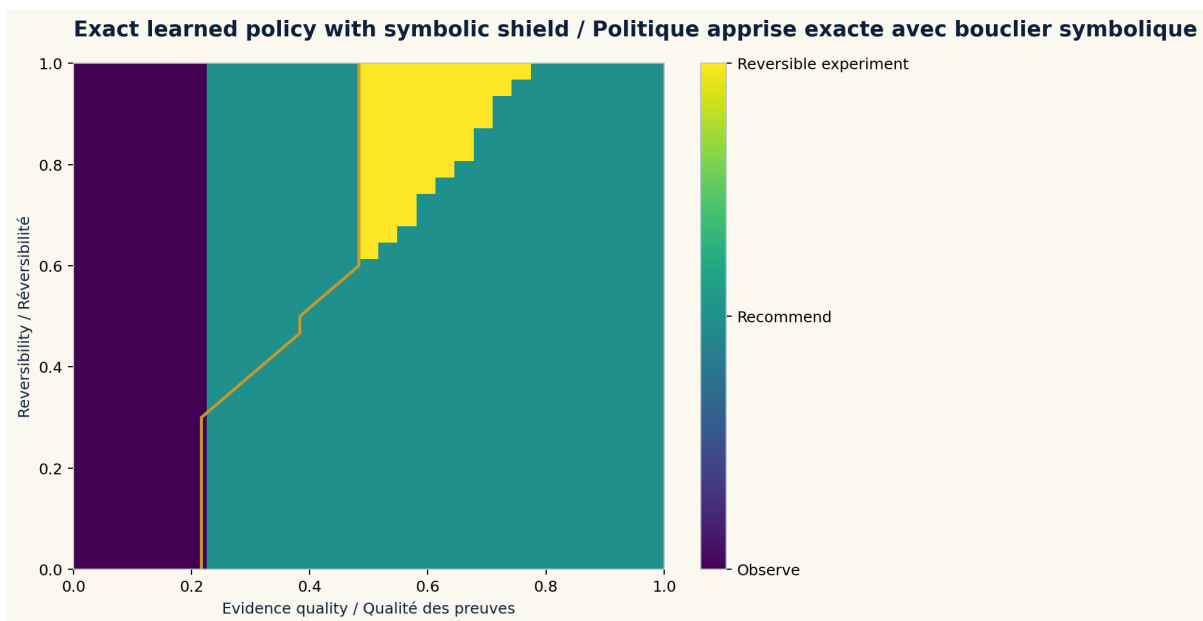


FIGURE 12.1. – Atlas exact des décisions neuronales-symboliques dans la trousse de simulation.

12.2 Exploitation A0 authentifiée

La simulation technique sous-jacente valide 34 tests du moteur sur 34, huit sondes d'inférence matériellement différentes, le jeu idempotent exact, le rejet des conflits, le rejet de l'altération, les dossiers

d’audit à ajout seulement et une fermeture propre. Les effets externes demeurent nuls. La simulation de pointe réutilise cette liaison exacte des preuves plutôt que de prétendre à un nouveau déploiement de production.

TABLE 12.1. – Décisions bornées représentatives.

Condition	Proposition neuronale	Résultat régi
Preuves fortes, grande réversibilité, faible risque pour les parties prenantes	Expérience réversible	Proposition d’expérience réversible, A0 seulement
Anomalie forte, faible réversibilité	Expérience réversible	Recommandation seulement
Preuves faibles	Recommander ou observer	Observer
Dérive importante	Toute proposition	Observer et figer
Risque élevé pour les parties prenantes	Expérience ou recommandation	Recommandation seulement

12.3 Successeur opérationnel R3 et filiation immuable de l’étude

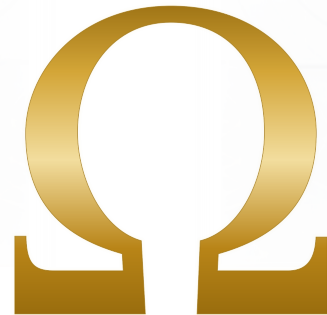
Les résultats scientifiques exacts demeurent liés à l’artéfact R2 figé et à l’identité du candidat. La trousse acheteur R3 publiée par la suite corrige à la source le routage de la Console opérateur, redirige ‘/console’ vers ‘/console/’, sert les ressources de même origine à partir de chemins canoniques et ajoute une acceptation dans un véritable navigateur. Cette correction opérationnelle ne réécrit ni les preuves R2, ni les octets du candidat, ni le résultat d’Alpha local, ni l’arithmétique du portefeuille.

Constat de version résolu

Utiliser R3 pour la livraison actuelle aux clients et pour toute nouvelle évaluation opérationnelle. Conserver R2 uniquement comme intrant de recherche immuable de la présente étude. Toute preuve indépendante doit préengager la version exacte et les octets du candidat évalués ; la succession opérationnelle ne modifie jamais silencieusement l’objet soumis à la preuve.

Équilibres stratégiques et thermodynamique institutionnelle

Incitatifs des rôles, jeux de preuve, modèles statistico-physiques, contrôle hamiltonien, énergie libre de Gibbs et accessibilité sûre.



LA VALEUR PEUT JUSTIFIER LA PREUVE; ELLE N'ACHÈTE JAMAIS LA PREUVE, L'ADMISSION OU L'AUTORITÉ.

13 Gouvernance institutionnelle et droits décisionnels

13.1 Séparation des rôles

Une preuve crédible exige des rôles distincts :

Demandeur

détient l'allégation de mission et le candidat, mais ne peut les certifier.

Gardien des preuves

contrôle les nouvelles preuves et empêche les fuites vers l'entraînement ou l'ajustement des seuils.

Vérificateur

évalue l'allégation préengagée et signe le résultat de preuve.

Autorité d'admission

décide si un candidat démontré peut entrer dans un environnement client.

Autorité opérationnelle

détermine les droits d'action effectifs maximaux et peut demeurer inférieure au plafond d'admission.

Une même organisation ne peut regrouper plusieurs rôles que si le niveau d'indépendance et les conflits sont divulgués puis acceptés. La preuve visée doit employer un vérificateur distinct du demandeur, de l'entraîneur et de l'exploitant.

13.2 Objets décisionnels

L'architecture institutionnelle traite chaque transition conséquente comme un dossier explicite : figement du candidat, plan de preuve, reçu de preuve, admission, plan de canari, décision de passerelle, dossier pré-action du contrôleur, dossier post-action, résultat du canari, attestation de portabilité, renouvellement ou succession. Ainsi, un score élevé du modèle ne peut se transformer silencieusement en autorité opérationnelle.

13.3 Contrôles du conseil

TABLE 13.1. – Contrôles minimaux du conseil avant toute décision d'admission.

Contrôle	Condition requise
Identité exacte	Empreintes du candidat, du comparateur, de la métrique et du plan de preuve figées
Nouvelles preuves	Aucun chevauchement avec l'entraînement, la sélection ou l'ajustement des seuils
Indépendance	Gardien et vérificateur distincts du demandeur et de l'exploitant
Dénominateur complet	C, R, H, I, D inclus et versionnés
Preuves négatives	ÉCHEC, abstention, dérive et événements défavorables conservés
Autorité	A0 pendant toute la preuve ; aucun effet externe
Admission	Décision responsable distincte après la preuve
Retour arrière	Mis à l'essai avant tout droit d'action ultérieur
Succession	Toute modification importante crée une nouvelle version et une nouvelle obligation de preuve

14 Gouvernance stratégique et physique et conception du mécanisme

14.1 Finalité et interprétation

Les chapitres précédents établissent une filiation exacte des artefacts, un candidat qualifié localement, un modèle transparent de valeur et une limite d'autorité A0. Le présent chapitre ajoute une couche décisionnelle formelle issue de la théorie des jeux, de la physique statistique, de la dynamique hamiltonienne et de l'énergie libre de Gibbs. Il ne s'agit pas d'une analogie décorative. L'objectif est de rendre explicites quatre questions institutionnelles :

1. Quelles incitations rendent la preuve véridique stratégiquement stable ?
2. Comment une institution doit-elle raisonner sur un grand ensemble d'états de candidats et de gouvernance sous incertitude ?
3. Quelles parties de la formation du candidat peuvent demeurer réversibles, et où les contrôles institutionnels irréversibles doivent-ils apparaître ?
4. Comment combiner valeur, risque, déficit de preuve et exposition d'autorité sans permettre à un candidat de grande valeur d'acheter sa propre preuve ou sa propre autorité ?

Limite d'interprétation

Les variables appelées température, entropie, Hamiltonien et énergie libre sont des constructions décisionnelles mesurées dans des unités institutionnelles déclarées. Elles ne désignent ni température physique, ni entropie thermodynamique, ni joules, et n'affirment pas qu'une organisation est littéralement un système moléculaire. Leur valeur est structurelle : elles rendent visibles les arbitrages, la métastabilité, les barrières et les transitions admissibles tout en préservant la distinction absolue entre capacité, preuve, admission et autorité.



FIGURE 14.1. – Synthèse stratégique et physique. Les quatre perspectives analytiques contraignent le même candidat sous le contrôle du client, tandis que l'autorité effective demeure A0.

14.2 Le jeu institutionnel de la preuve

Soit l'ensemble de rôles

$$\mathcal{N} = \{c, v, d, o\}, \quad (14.1)$$

où c est le demandeur, v le vérificateur indépendant, d l'organe d'admission et o l'exploitant. Chaque rôle agit sur un objet différent et signe au moyen d'un justificatif distinct. Le demandeur peut proposer un candidat et une thèse de valeur, mais ne peut émettre un reçu de preuve. Le vérificateur peut accepter ou refuser l'allégation de preuve, mais ne peut accorder l'admission en production. L'organe d'admission peut accorder un statut de déploiement, mais ne peut élargir silencieusement l'enveloppe d'autorité opérationnelle. L'exploitant ne peut agir que dans l'autorité effective constituée séparément.

Pour le rôle i , écrivons une utilité stylisée :

$$u_i(a) = B_i(a) - K_i(a) - \Pi_i \mathbf{1}\{m_i \wedge \delta_i\} - L_i^{\text{access}} \mathbf{1}\{m_i \wedge \delta_i\} - L_i^{\text{reputation}} \mathbf{1}\{m_i \wedge \delta_i\}, \quad (14.2)$$

où B_i est le bénéfice légitime, K_i le coût d'exécution du rôle, m_i une manipulation, δ_i sa détection, Π_i une sanction directe et les termes L_i les pertes d'accès et de réputation. Une condition suffisante pour que l'action véridique constitue une meilleure réponse est

$$g_i^{\text{manip}} \leq \Pr(\delta_i \mid m_i) \Pi_i + L_i^{\text{access}} + L_i^{\text{reputation}}, \quad (14.3)$$

où g_i^{manip} est le gain privé maximal associé à la manipulation.

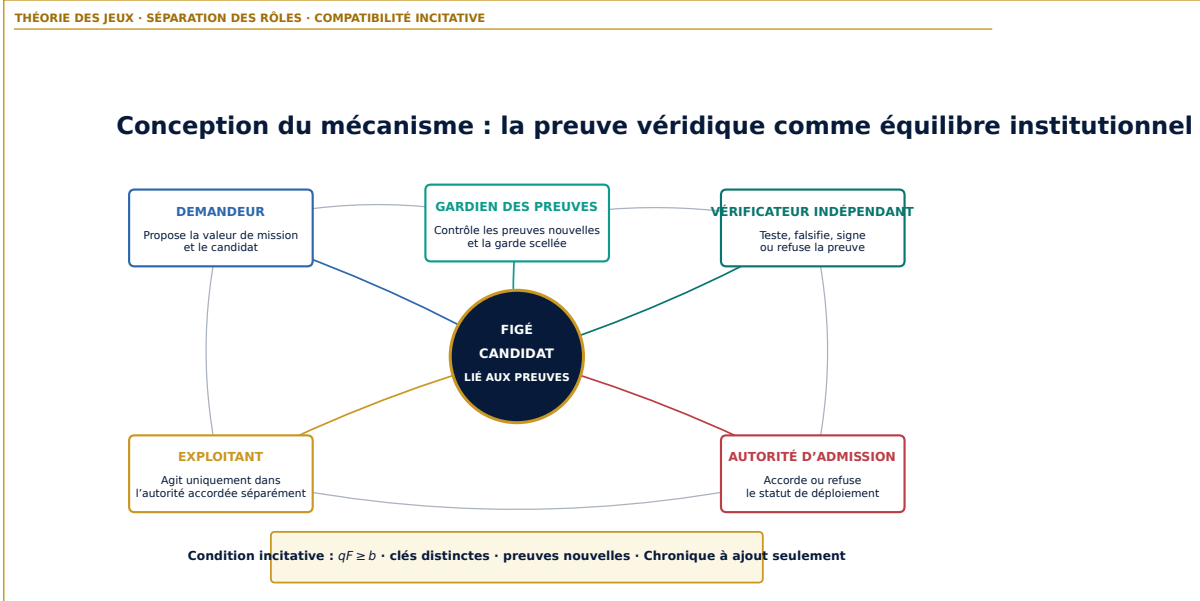


FIGURE 14.2. – Conception du mécanisme du jeu de preuve. La séparation des rôles, les preuves nouvelles, les clés distinctes, le préengagement et une Chronique à ajout seulement augmentent le coût attendu de la manipulation et réduisent le gain de collusion.

L'architecture n'affirme pas que l'inégalité de l'équation (14.3) a été estimée empiriquement pour une institution réelle. Il s'agit d'une obligation de conception. Avant toute utilisation opérationnelle, l'institution doit préciser les probabilités de détection, les sanctions directes, la révocation d'accès, les conflits d'intérêts et le modèle de rémunération admissible. Cette exigence suit la logique de l'équilibre de Nash et de la conception de mécanismes : le résultat souhaité doit être stable sous les incitations des rôles participants, et non seulement sous leur bonne volonté (NASH, 1950 ; MYERSON, 1981).

14.2.1 Représentation comme jeu potentiel

Un potentiel institutionnel utile est

$$\Phi(a) = V_{\text{mission}}(a) - \lambda_R R(a) - \lambda_E E_{\text{deficit}}(a) - \lambda_C C_{\text{complete}}(a) - \lambda_X X_{\text{role collision}}(a) - \lambda_A A_{\text{excess}}(a). \quad (14.4)$$

Si les changements unilatéraux de rôles satisfont

$$u_i(a'_i, a_{-i}) - u_i(a_i, a_{-i}) = \Phi(a'_i, a_{-i}) - \Phi(a_i, a_{-i}), \quad (14.5)$$

alors le mécanisme de preuve constitue un jeu potentiel exact au sens de MONDERER et SHAPLEY (1996). La conséquence pratique est directe : les récompenses et sanctions doivent orienter l'incitation locale de chaque rôle dans le même sens que la valeur de mission de l'institution après déduction du risque, du déficit de preuve, des conflits et de l'autorité excessive.

Proposition 14.1 (Absence d'autocertification). Si le demandeur ne peut écrire un reçu de vérification, si les preuves de démonstration sont nouvelles et placées sous contrôle indépendant et si le justificatif du vérificateur n'est pas contrôlé par le demandeur, alors la formation du candidat ne peut à elle seule faire passer P de la qualification locale à la preuve indépendante.

Démonstration. La relation de transition autorisée ne contient aucune arête signée par le demandeur entre P_{local} et $P_{\text{independent}}$. Toute tentative d'arête ne comporte ni le justificatif requis du vérificateur ni la liaison aux preuves nouvelles; elle se situe donc à l'extérieur de l'ensemble des transitions admissibles. $\square$

14.3 Microétats institutionnels et ensemble de Gibbs

Définissons un microétat institutionnel

$$x = (\theta, z, e, r, S), \quad (14.6)$$

où θ représente les paramètres appris, z les programmes symboliques, e l'état des preuves, r la configuration des rôles et justificatifs, et $S = (C, P, D, A)$ l'état institutionnel. Soit $\mathcal{X}_{\text{adm}}$ l'ensemble des microétats qui satisfont chaque règle absolue de la constitution de mission.

Un Hamiltonien institutionnel ajusté au risque peut s'écrire

$$\mathcal{H}(x) = C(x) + R(x) + H(x) + I(x) + D(x) + \lambda_E E_{\text{deficit}}(x) + \lambda_A A_{\text{excess}}(x) + \lambda_X X_{\text{conflict}}(x) - \rho V_{\text{mission}}(x), \quad (14.7)$$

avec $\mathcal{H}(x) = +\infty$ pour toute violation d'une règle absolue. Cette convention est décisive : aucune valeur projetée ne peut compenser un état d'autorité inadmissible, un rôle de preuve manquant ou une réutilisation interdite des preuves.

Pour une température de gouvernance $\tau > 0$, définissons

$$Z(\tau) = \sum_{x \in \mathcal{X}_{\text{adm}}} \exp \left[-\frac{\mathcal{H}(x)}{\tau} \right], \quad (14.8)$$

$$\pi_\tau(x) = \frac{1}{Z(\tau)} \exp \left[-\frac{\mathcal{H}(x)}{\tau} \right], \quad (14.9)$$

$$\mathcal{G}_\tau = -\tau \log Z(\tau). \quad (14.10)$$

Ici, τ représente l'échelle déclarée d'incertitude, de volatilité et d'adversité institutionnelles. Lorsque τ est faible, la distribution se concentre sur les états admissibles à faible Hamiltonien. Lorsque τ augmente, l'institution préserve davantage de diversité, car une concentration prématurée devient coûteuse. Il s'agit de la logique d'entropie maximale associée aux ensembles de Gibbs et à la formulation informationnelle de la mécanique statistique de Jaynes (JAYNES, 1957; PATHRIA et BEALE, 2011).

La formulation variationnelle équivalente est

$$\mathcal{G}_\tau = \min_{q \in \Delta(\mathcal{X}_{\text{adm}})} \left\{ \mathbb{E}_q[\mathcal{H}(x)] - \tau H(q) \right\}, \quad (14.11)$$

où $H(q) = -\sum_x q(x) \log q(x)$ est l'entropie de Shannon (SHANNON, 1948). L'institution minimise donc le coût attendu ajusté au risque tout en conservant assez d'entropie pour éviter une convergence prématurée et fragile sous incertitude.

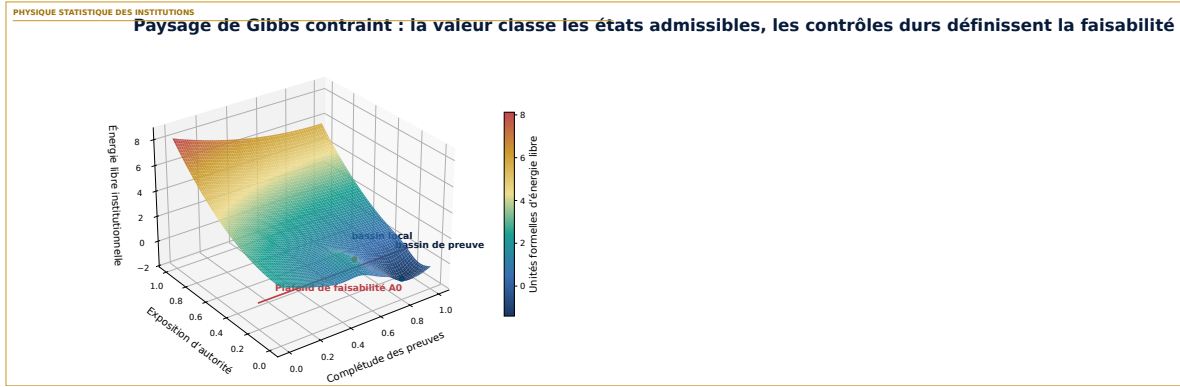


FIGURE 14.3. – Paysage décisionnel d'énergie libre de Gibbs. La qualification locale peut constituer un minimum métastable ; une barrière de vérification sur preuves nouvelles la sépare de la preuve indépendante. Les règles absolues de preuve et d'autorité demeurent discrètes.

Interprétation institutionnelle de l'énergie libre de Gibbs

Une énergie G_τ plus faible ne constitue pas un certificat. Elle désigne un état admissible attrayant selon les hypothèses déclarées. Une allégation de preuve exige toujours le protocole exact de preuve. Une admission exige toujours l'organe d'admission. L'autorité exige toujours une décision distincte d'autorité bornée. La minimisation de l'énergie libre peut classer les états faisables ; elle ne peut légaliser un état inadmissible.

14.4 Formation hamiltonienne avant la frontière de figement

La formation du candidat peut être représentée dans un espace des phases augmenté comportant des coordonnées q et une impulsion p :

$$\mathcal{H}(q, p) = \frac{1}{2} p^\top M^{-1} p + U(q), \quad (14.12)$$

avec la dynamique canonique

$$\dot{q} = M^{-1} p, \quad \dot{p} = -\nabla U(q). \quad (14.13)$$

Le potentiel U peut coder la valeur négative de la mission, augmentée de pénalités relatives aux violations de sécurité, à la faiblesse des preuves, à la complexité, à la fragilité et au risque d'imitation. L'intégration symplectique préserve la structure qualitative du flot de phase et devient donc utile lorsque l'exploration réversible et la fidélité de la dynamique à long terme comptent (HAIRER et al., 2006). Des formulations connexes du contrôle emploient des transformations exponentielles et des structures d'intégrales de chemin afin de relier contrôle stochastique, coût informationnel et mesures de Gibbs (KAPPEN, 2005 ; TODOROV, 2009 ; ORTEGA et BRAUN, 2013).

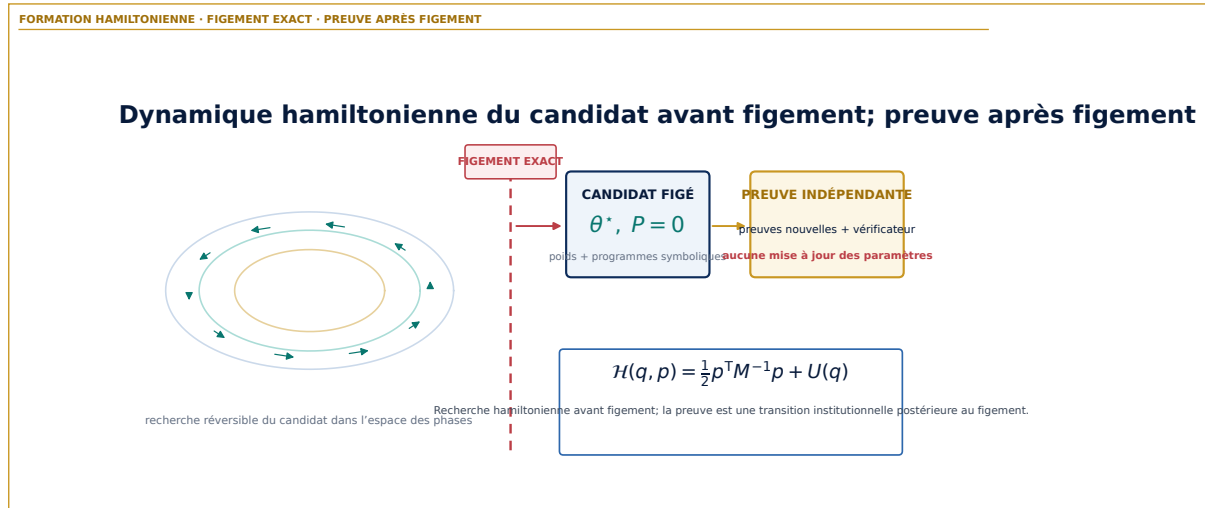


FIGURE 14.4. – Formation hamiltonienne et discontinuités institutionnelles. La recherche réversible de candidats a lieu avant le figement exact. La preuve indépendante évalue l’objet figé sur des preuves nouvelles sans effectuer de mise à jour des paramètres.

Le candidat R2 mis à l’essai n’a pas été entraîné au moyen de l’intégrateur hamiltonien de l’équation (14.12). Le formalisme constitue une extension architecturale et un test de conception : tout moteur futur de formation peut être remplacé, mais le figement doit fixer $(q, p) = (q^*, 0)$ pour l’objet évalué, et l’étape de preuve ne doit pas utiliser les preuves de démonstration pour modifier q^* .

Proposition 14.2 (Barrière de preuve). Si le protocole de preuve interdit les mises à jour de paramètres après le figement du candidat et lie chaque résultat à l’empreinte exacte de q^* et z^* , alors la transition de preuve ne constitue pas une poursuite de la dynamique d’entraînement. Elle constitue un essai institutionnel discret appliqué à un objet fixe.

14.5 Température de gouvernance et transitions de phase

Soit $p \in [0, 1]$ la complétude de la preuve et τ la température de gouvernance. Quatre régions qualitatives sont utiles :

1. **Formation / exploration** : faible complétude de la preuve ; la diversité des candidats possède une grande valeur.
2. **Candidat prêt pour la preuve** : la qualification locale est terminée, l’objet est figé et un contrat de preuve existe.
3. **Démontré indépendamment** : les preuves nouvelles et un reçu valide du vérificateur satisfont à la règle préengagée.
4. **Arrêter / requalifier** : l’incertitude ou l’adversité est assez élevée pour que les preuves existantes ne permettent plus de progresser.

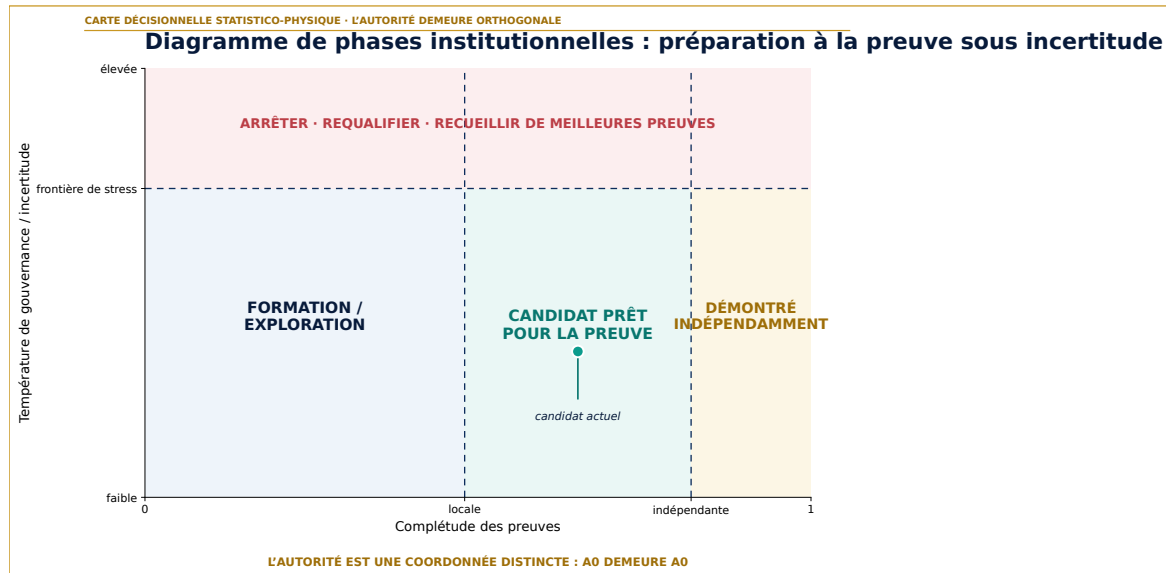


FIGURE 14.5. – Diagramme de phases institutionnelles. Selon les hypothèses de simulation, le candidat actuel se trouve dans la région prête pour la preuve. Une hausse de l'incertitude peut ramener l'institution à la formation ou imposer l'arrêt. L'autorité demeure une coordonnée orthogonale.

La frontière de phase n'est pas universelle. Chaque constitution de mission doit définir les paramètres d'ordre, les seuils, l'échelle d'incertitude et les contrôles absolus. L'invariant important est qu'une transition de phase de preuve ne produit pas de transition de phase d'autorité.

Proposition 14.3 (Aucune autorité gratuite). Supposons que l'ensemble d'états admissibles attribue un Hamiltonien infini à tout état $A > A_0$ qui ne comporte pas une décision valide d'autorité bornée. Alors ni la maximisation de valeur ni la minimisation de l'énergie libre de Gibbs ne peuvent sélectionner $A > A_0$ en l'absence de cette décision.

Démonstration. Tout état d'autorité supérieure non autorisé possède un poids de Gibbs nul, puisque $\exp(-\infty/\tau) = 0$, et se trouve exclu de l'optimiseur faisable. La conclusion demeure vraie quelle que soit la valeur de mission projetée. $\square$

14.6 Relier l'Alpha de mission au budget d'énergie libre

Le modèle d'Alpha à dénominateur complet soustrait déjà le coût complet, la réserve de risque, la charge humaine, la réserve d'imitation et la réserve de dérive. En notation stratégique et physique, le calcul central de souscription correspond à la partie finie de

$$\mathcal{H}_{\text{value}} = (C + R + H + I + D) - \rho G, \quad (14.14)$$

de sorte que

$$\alpha_{\text{portfolio}} = -\mathcal{H}_{\text{value}}. \quad (14.15)$$

Pour le scénario central déclaré, $\rho G = 321,4872$ M \$ CA et les déductions complètes totalisent 103 M \$ CA, ce qui produit $\alpha_{\text{portfolio}} = 218,4872$ M \$ CA. Le scénario inférieur commence par la borne inférieure locale de sensibilité et produit 70,5488 M \$ CA.

Cette identité ne transforme pas l'Alpha de scénario en observable thermodynamique. Elle rend la comptabilité transparente : la valeur de mission abaisse l'Hamiltonien institutionnel ; les coûts et les risques l'augmentent. Les déficits de preuve, les conflits de rôles et l'autorité excessive constituent des pénalités supplémentaires, tandis que les violations absolues reçoivent un coût infini.

TABLE 14.1. – Correspondance entre physique stratégique, théorie des jeux et gouvernance institutionnelle.

Objet analytique	Signification institutionnelle	Conséquence opérationnelle
Profil stratégique a	Actions signées du demandeur, du vérificateur, de l'organe d'admission et de l'exploitant	Évaluer les incitations et les voies de collusion
Potentiel $\Phi(a)$	Valeur de mission après pénalités de risque, de preuve, de conflit et d'autorité	Accepter les changements unilatéraux uniquement lorsqu'ils améliorent la valeur institutionnelle
Microétat x	Candidat exact, preuves, rôles et état CPDA	Hacher, versionner et auditer l'état complet
Hamiltonien $\mathcal{H}(x)$	Coût et risque complets moins valeur de mission	Classer les états admissibles ; rejeter les violations absolues
Température τ	Volatilité, incertitude et adversité	Préserver plus de diversité et exiger des preuves plus fortes lorsque τ augmente
Entropie $H(q)$	Diversité des hypothèses ou politiques admissibles	Éviter une convergence prématurée et fragile
Énergie libre $\mathcal{G}_\tau$	Valeur ajustée au risque avec arbitrage entre incertitude et diversité	Sélectionner le meilleur état admissible, jamais un état non autorisé
Frontière de figement	Identité immuable du candidat exact	Séparer les preuves d'entraînement des preuves de démonstration
Transition de phase	Changement valide d'état de preuve ou d'admission	Exiger une décision signée propre au contrôle

14.7 Protocole d'exploitation stratégique et physique

Une institution à haute assurance peut appliquer le formalisme au moyen du protocole suivant :

1. **Constituer le jeu.** Nommer le responsable de mission, le demandeur, le gardien des preuves, le vérificateur, l'organe d'admission et l'exploitant ; divulguer les conflits et le contrôle des clés.
2. **Définir l'espace des états faisables.** Encoder les contraintes de preuve, de sécurité, de réversibilité, de confidentialité, d'autorité et de séparation des rôles. Attribuer un coût infini aux états inadmissibles.
3. **Former un ensemble diversifié de candidats.** Utiliser la diversité des modèles, des méthodes de diversité-qualité ou une recherche régularisée par l'entropie plutôt qu'une trajectoire unique non contestée.
4. **Falsifier et figer.** Retirer les hypothèses faibles, lier le modèle et les programmes symboliques sélectionnés, fixer l'empreinte de figement et arrêter les mises à jour de paramètres.

5. **Calculer l'Alpha local de mission.** Comparer au meilleur référentiel figé au moyen de déductions à dénominateur complet et d'une incertitude explicite.
6. **Fixer la température de gouvernance.** Augmenter τ lorsque la dérive, l'adversité, la latence des preuves ou l'incertitude du modèle s'accroît ; élargir l'exploration ou arrêter en conséquence.
7. **Jouer le jeu de preuve indépendante.** Employer des preuves nouvelles, des justificatifs distincts, des métriques préengagées et une Chronique à ajout seulement.
8. **Décider séparément la preuve, l'admission et l'autorité.** Une preuve favorable ne modifie que P . Des décisions ultérieures peuvent modifier D et A selon leurs propres constitutions.

Règle de décision exécutive

Financer l'état suivant uniquement lorsqu'il est à la fois *économiquement attrayant* et *institutionnellement admissible*. En notation formelle, sélectionner

$$x^* \in \arg \min_{x \in \mathcal{X}_{\text{adm}}} \mathcal{G}_\tau(x),$$

mais autoriser une transition uniquement lorsque le contrôle désigné la signe. L'optimisation classe les candidats ; les institutions accordent les statuts et l'autorité.

14.8 Limites du formalisme stratégique et physique

Le cadre demeure rigoureux quant à ses propres limites. Les correspondances, pondérations, températures, pénalités et espaces d'états sont propres à la mission et doivent être étalonnés empiriquement. Les minima d'énergie libre peuvent être erronés lorsque le modèle est mal spécifié. L'alignement d'un jeu potentiel peut échouer lorsque les gains privés sont cachés. Des rôles en collusion peuvent contourner une séparation nominale si le contrôle des justificatifs n'est pas réellement indépendant. Les conditions hors équilibre peuvent invalider les interprétations stationnaires ; les relations de fluctuation rappellent l'importance des transitions rares et de la dépendance au chemin (JARZYNSKI, 1997 ; CROOKS, 1999). Aucune élégance mathématique ne remplace les preuves nouvelles, les personnes responsables, l'autorité juridique, les exercices opérationnels ou le droit de refuser le déploiement.

15 Synthèse de pointe : équilibres stratégiques, thermodynamique de l'information et accessibilité sûre

15.1 Objet : un calcul décisionnel, non une métaphore

Le chapitre stratégique et physique précédent a introduit des rôles de preuve compatibles avec les incitatifs, un ensemble de Gibbs contraint et la formation hamiltonienne du candidat. Le présent chapitre referme la boucle mathématique. Il traite l'acquisition des preuves comme un gain d'information, le comportement des vérificateurs comme un jeu évolutif, la progression opérationnelle comme un problème d'accessibilité et l'escalade prématurée comme un événement de grande déviation. L'objectif est un calcul décisionnel institutionnel qui rende les conditions de refus aussi explicites que les conditions de valeur.

Statut formel

Les constructions ci-dessous sont normatives et décisionnelles. Elles organisent les preuves, les incitatifs, l'incertitude et les transitions admissibles. Elles ne transforment pas une étude synthétique en mesure physique, preuve indépendante, admission en production ou valeur réalisée. Les règles dures de la mission conservent la priorité sur tout objectif scalaire.

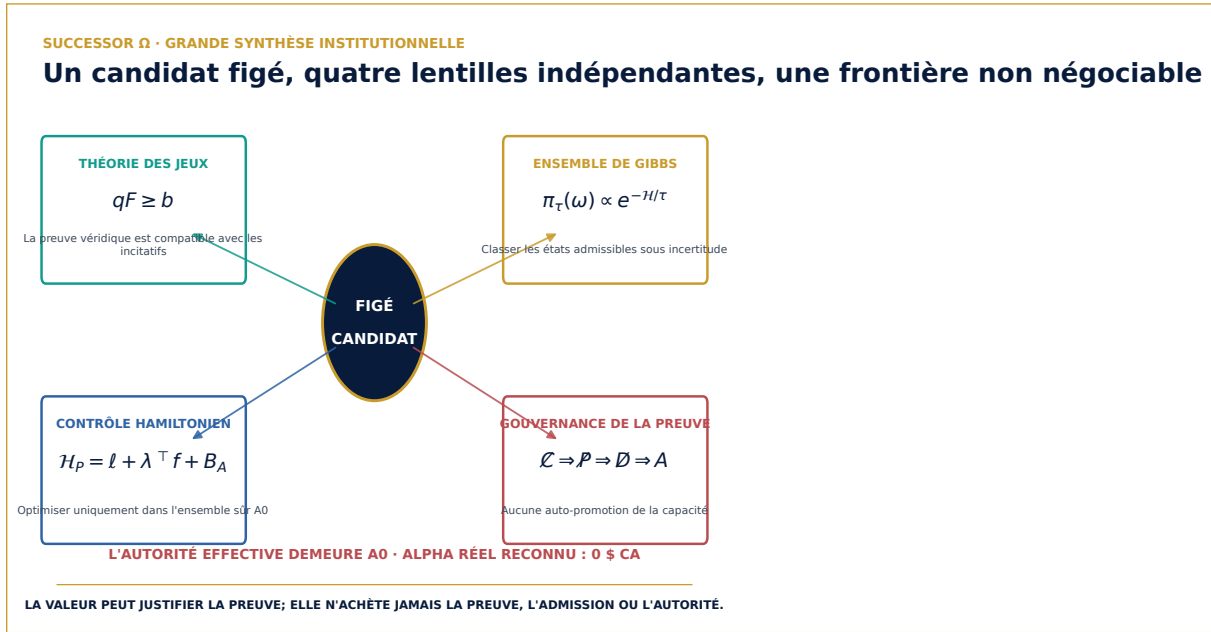


FIGURE 15.1. – Grande synthèse. Quatre lentilles analytiques agissent sur le même candidat figé exact tandis que la frontière des allégations et l'autorité A0 demeurent invariantes.

15.2 Les preuves comme gain d'information

Soit H_1 la proposition selon laquelle le candidat figé exact surpasse le comparateur authentifié le plus fort selon le contrat de preuve, et soit H_0 son complément. Pour les preuves nouvelles $E_{1:t}$, les rapports de probabilités a posteriori satisfont

$$\log \frac{\Pr(H_1 | E_{1:t})}{\Pr(H_0 | E_{1:t})} = \log \frac{\Pr(H_1)}{\Pr(H_0)} + \sum_{s=1}^t \log \Lambda(E_s), \quad \Lambda(E_s) = \frac{p(E_s | H_1)}{p(E_s | H_0)}. \quad (15.1)$$

Un programme de preuve n'est donc pas l'accumulation d'anecdotes convaincantes. C'est un processus contrôlé d'acquisition de preuves fondées sur les rapports de vraisemblance, selon des règles préengagées d'échantillonnage, de garde, de comparaison, de mesure et d'arrêt.

Si $p_t(h)$ est la distribution soutenue par l'institution sur les hypothèses admissibles, son entropie de soutien est

$$S_t = - \sum_h p_t(h) \log p_t(h). \quad (15.2)$$

Le gain d'information attendu d'une action probante a peut s'écrire

$$\mathcal{I}(a) = \mathbb{E}_{E \sim p(\cdot|a)} [D_{\text{KL}}(p(h | E, a) \| p(h))]. \quad (15.3)$$

L'institution devrait privilégier les actions probantes qui maximisent le gain d'information pertinent par unité complète de coût et de charge, sous réserve des contraintes d'indépendance et d'admissibilité (SHANNON, 1948; COVER et THOMAS, 2006; KULLBACK et LEIBLER, 1951).

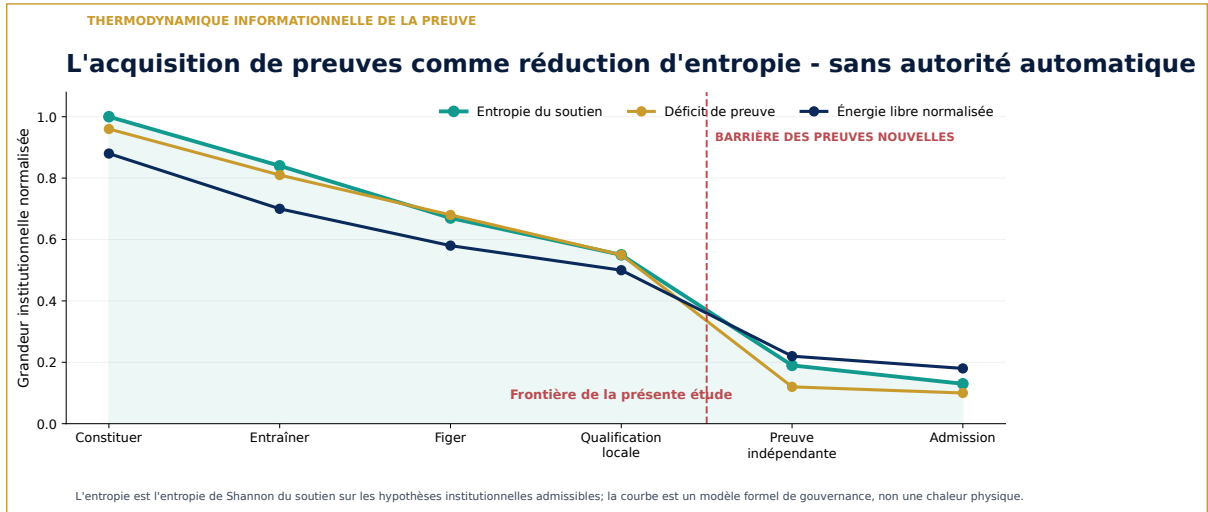


FIGURE 15.2. – Profil informationnel et thermodynamique de la preuve. La qualification locale réduit l'incertitude, mais la contraction probante la plus importante survient uniquement après le franchissement de la barrière des preuves nouvelles.

Proposition 15.1 (Monotonie des preuves sous mises à jour valides). Supposons que chaque élément de preuve admis soit nouveau, sous contrôle indépendant et intégré au moyen d'une mise à jour de vraisemblance appropriée. L'entropie a posteriori attendue ne peut alors augmenter en espérance. Cette réduction n'implique ni admission ni autorité, car ces coordonnées obéissent à des règles de transition distinctes.

Démonstration. L'entropie conditionnelle n'est pas supérieure à l'entropie antérieure en espérance ; de manière équivalente, l'information mutuelle est non négative. La conclusion touche uniquement l'état informationnel de la preuve. Aucune règle de transition de l'entropie réduite vers l'admission ou l'autorité n'est supposée. $\square$

15.3 Valeur sensible au risque et énergie libre de style Gibbs

Un critère de valeur attendue neutre au risque peut survaloriser les états dont la moyenne est attrayante, mais dont les queues sont inacceptables. Soit $L(x)$ la perte de mission à dénominateur complet pour l'état admissible x . La fonctionnelle de risque entropique est

$$\rho_\tau(L) = \tau \log \mathbb{E} \left[\exp \left(\frac{L}{\tau} \right) \right], \quad \tau > 0, \quad (15.4)$$

qui tend vers la perte attendue lorsque la sensibilité au risque s'affaiblit et accorde un poids croissant aux queues défavorables lorsque cette sensibilité augmente. Un objectif correspondant d'énergie libre institutionnelle est

$$\mathcal{F}_\tau(q) = \mathbb{E}_q[\mathcal{H}(x)] - \tau \mathcal{H}(q), \quad q^* \in \arg \min_{q \in \Delta(\mathcal{X}_{\text{adm}})} \mathcal{F}_\tau(q). \quad (15.5)$$

Le premier terme favorise les états de faible Hamiltonien ajusté au risque. Le second préserve une diversité stratégique lorsque les preuves sont incomplètes. Le minimiseur suit une distribution de Gibbs sur l'ensemble des états admissibles ; les états de preuve ou d'autorité infaisables reçoivent une masse nulle parce que leur Hamiltonien est infini.

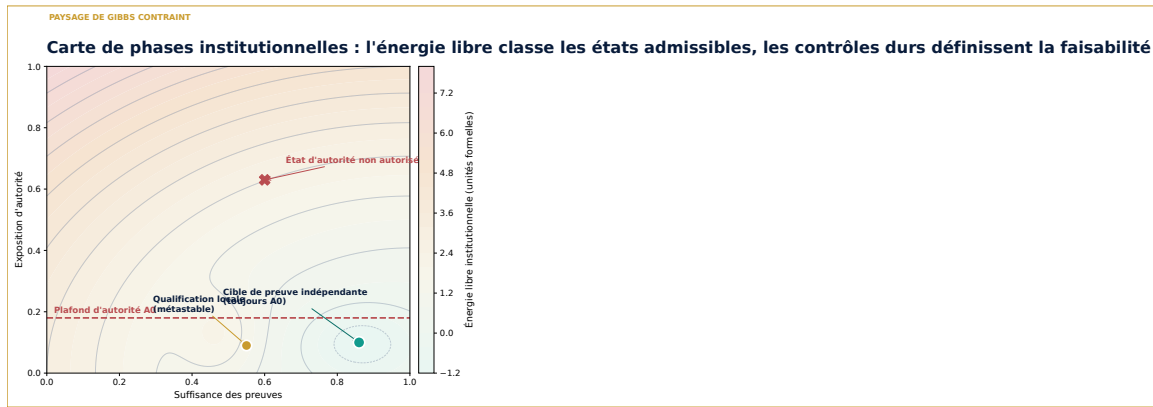


FIGURE 15.3. – Carte de phases contrainte. L'énergie libre classe les états admissibles ; les contrôles durs de preuve et d'autorité déterminent les états qui appartiennent à l'ensemble faisable.

TABLE 15.1. – Signification décisionnelle des grandeurs stratégiques et physiques.

Grandeur	Interprétation institutionnelle	Ce qu'elle ne peut pas faire
$\mathcal{H}(x)$	Score complet de valeur, risque, preuve et autorité d'un microétat admissible	Déroger à une règle dure
τ	Échelle déclarée d'incertitude, de volatilité et d'adversarialité	Mesurer une température physique
$\mathcal{H}(q)$	Entropie de soutien sur les hypothèses ou états admissibles	Certifier une allégation
$\mathcal{F}_\tau$	Classement sensible au risque des distributions d'états faisables	Accorder une preuve, une admission ou une autorité
$Z(\tau)$	Normalisateur sur les états admissibles	Créer des états exclus par la constitution

15.4 Écologie évolutive des vérificateurs

Un vérificateur peut être honnête tandis que l'institution de vérification demeure fragile. Soit $x \in [0, 1]$ la part de la capacité de vérification utilisant une stratégie véridique et soit $q \in [0, 1]$ l'intensité effective de l'audit. Une dynamique de réplication stylisée est

$$\dot{x} = x(1 - x)(\Delta u_{V,C}(x, q)), \quad (15.6)$$

où $\Delta u_{V,C}$ est l'avantage de rendement de la vérification véridique par rapport à la vérification collusive ou négligente. Si le bénéfice privé de la manipulation est b , la pénalité effective est F et la probabilité de détection est q , une condition élémentaire de dissuasion est

$$qF \geq b. \quad (15.7)$$

Une institution robuste va plus loin : elle réduit b par la séparation des rôles et la garde des preuves, augmente q par l'audit aléatoire et la rejouabilité, et accroît le coût durable de la manipulation par la révocation, la réputation et les conséquences d'accès.

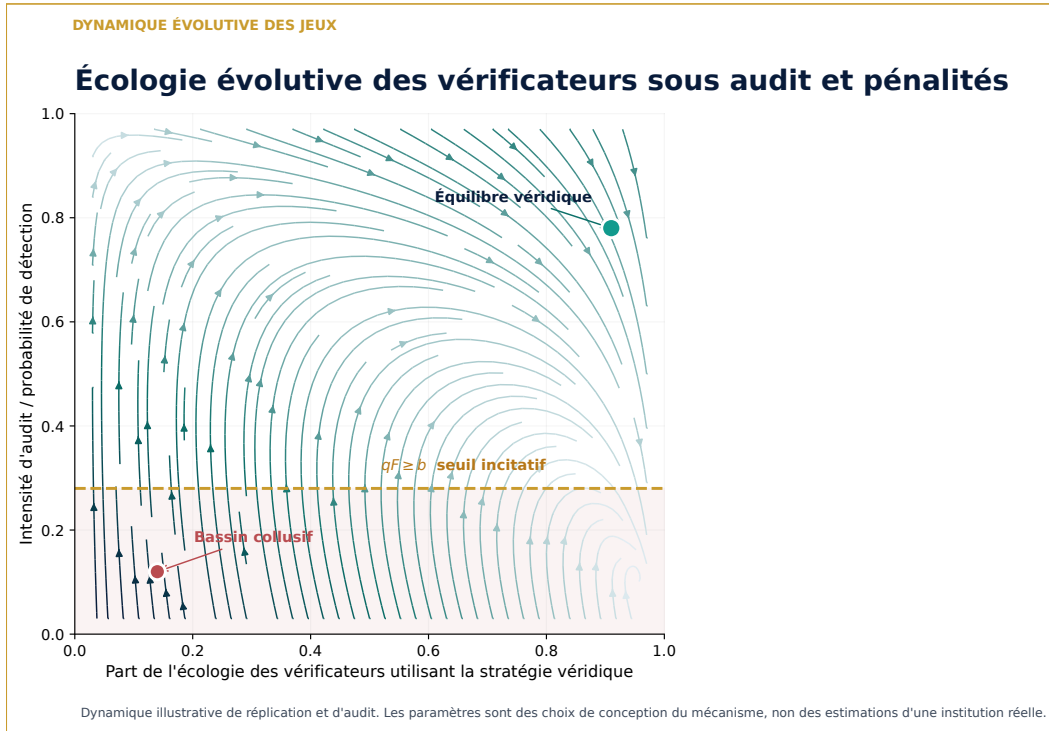


FIGURE 15.4. – Écologie évolutive illustrative des vérificateurs. Des incitatifs appropriés et une intensité d’audit suffisante élargissent le bassin d’attraction de la vérification véridique.

La figure n’est pas un modèle ajusté à un marché réel de la vérification. Elle révèle une obligation de conception : la preuve véridique devrait être une propriété d’équilibre du mécanisme, et non une hypothèse de personnalité concernant un seul examinateur (HOFBAUER et SIGMUND, 1998 ; BAŞAR et OLSDER, 1999).

Proposition 15.2 (La valeur n’achète pas la preuve). Que la valeur de mission projetée influence la décision de commander une preuve, mais non la rémunération du vérificateur conditionnellement au résultat. Si la rémunération du vérificateur est indépendante du résultat, si les preuves sont sous contrôle indépendant et si le coût attendu d’une fausse attestation est au moins égal à son bénéfice privé, une valeur projetée plus élevée ne rend pas à elle seule un faux succès optimal.

15.5 Accessibilité de Hamilton-Jacobi sous une enveloppe d’autorité dure

Soit x_t l’état contenant la suffisance des preuves, la dérive, l’exposition des parties prenantes, la réversibilité opérationnelle et l’état d’autorité. Soit u_t une commande institutionnelle proposée. L’ensemble sûr est

$$\mathcal{K} = \{x : h_j(x) \geq 0, j = 1, \dots, m\}, \quad (15.8)$$

où les h_j codent la constitution de mission, le contrôle de preuve, le contrôle d’admission et le plafond d’autorité. Une condition de fonction barrière exige une commande admissible u telle que

$$\nabla h_j(x)^\top f(x, u) + \alpha_j(h_j(x)) \geq 0 \quad \text{pour toute contrainte active } j. \quad (15.9)$$

La question d’accessibilité de Hamilton-Jacobi associée n’est pas « quel est le chemin le plus court vers une autorité supérieure ? ». Elle est « quelles trajectoires améliorant la preuve demeurent dans l’ensemble

sûr tandis que l'autorité reste fixe jusqu'à une transition autorisée distincte ? » (FLEMING et SONER, 2006 ; BERTSEKAS, 2017).

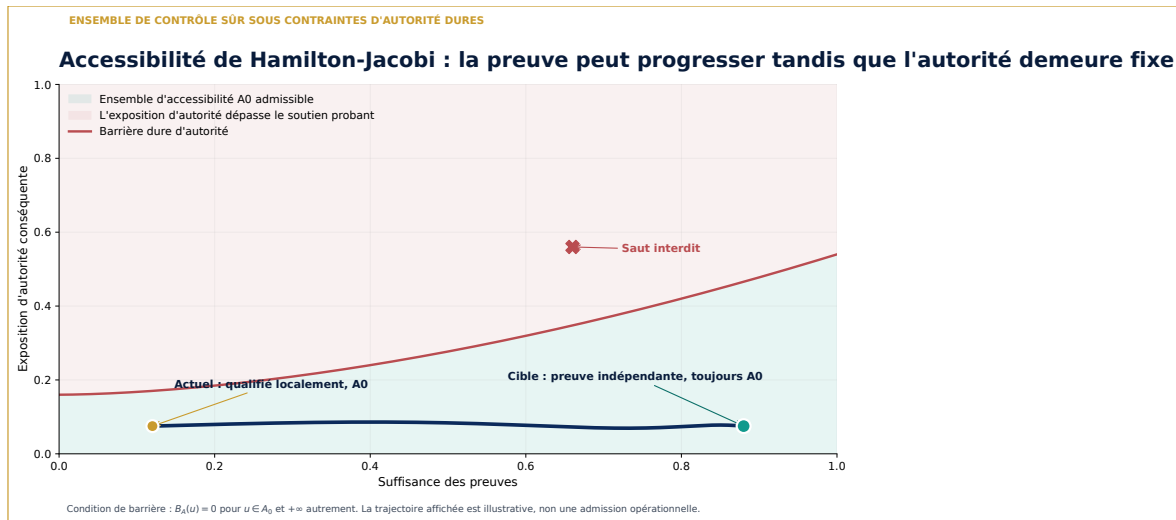


FIGURE 15.5. – Accessibilité sûre. Le candidat exact peut passer de la qualification locale vers la preuve indépendante sans passer à une coordonnée d'autorité supérieure.

Principe 15.1 (Non-covariance de l'autorité). Les trajectoires de preuve et de capacité ne sont pas tenues de covarier avec l'autorité. La politique sûre par défaut consiste à maintenir $A = A_0$ tandis que C et P évoluent, et à traiter toute modification de D ou de A comme un événement institutionnel signé distinct.

15.6 Frontière de refus fondée sur les grandes déviations

Les défaillances de gouvernance rares mais catastrophiques ne devraient pas disparaître dans le rendement moyen. Soit Y_n une statistique empirique de perte institutionnelle sur n cas nouveaux. Sous des conditions de régularité, la théorie des grandes déviations caractérise les probabilités de queue au moyen d'une fonction de taux I :

$$\Pr(Y_n \in F) \asymp \exp\{-n \inf_{y \in F} I(y)\}. \quad (15.10)$$

Pour un ensemble interdit F comprenant des actions non autorisées, des dommages critiques aux parties prenantes ou une rupture de la chaîne de preuves, l'institution devrait préengager un exposant minimal acceptable ou un contrôle direct de tolérance zéro. Cette exigence est plus stricte que la comparaison des moyennes. Elle demande si les états dangereux disparaissent assez rapidement avec les preuves ou demeurent structurellement accessibles (DEMBO et ZEITOUNI, 1998).

Une mesure complémentaire du risque de queue est la valeur à risque conditionnelle :

$$\text{CVaR}_\beta(L) = \min_{\eta} \left\{ \eta + \frac{1}{1-\beta} \mathbb{E}[(L - \eta)_+] \right\}. \quad (15.11)$$

Le contrat de preuve peut exiger à la fois une borne de confiance inférieure positive de l'Alpha à dénominateur complet et une CVaR_β bornée pour les interventions erronées, la charge des réviseurs et la perte de continuité (ROKAFELLAR et URYASEV, 2000).

15.7 Fonctionnelle conjointe d'admissibilité

Les couches mathématiques peuvent être combinées sans confondre les coordonnées institutionnelles. Définissons

$$\mathcal{J} = \underbrace{1\{\underline{\alpha} > h\}}_{\text{valeur}} \underbrace{1\{P = P_{\text{independent}}\}}_{\text{preuve}} \underbrace{1\{\Delta \mathcal{F}_\tau < 0\}}_{\text{stabilité sensible au risque}} \underbrace{1\{\text{CVaR}_\beta(L) \leq \ell_{\max}\}}_{\text{contrôle de queue}} \prod_{k=1}^K \underbrace{g_k}_{\text{contrôles durs}}. \quad (15.12)$$

Si $\mathcal{J} = 0$, l'admission est inadmissible. Si $\mathcal{J} = 1$, l'admission devient uniquement admissible à une décision responsable distincte. Aucun objectif scalaire, aucune probabilité a posteriori, aucun minimum d'énergie libre ni aucun équilibre de jeu ne signe le reçu d'admission.

TABLE 15.2. – Logique de refus destinée au conseil.

Condition	Interprétation	Action requise
Alpha local positif; preuve ouverte	Une thèse d'investissement existe, non une allégation de preuve	Commander des preuves nouvelles
Preuve indépendante réussie; admission ouverte	Supériorité établie pour la portée exacte	Décision de l'autorité d'admission
Admission réussie; autorité A0	Un état de déploiement existe, non un droit d'action	Maintenir A0 ou rendre une décision d'autorité distincte
Échec du contrôle de risque de queue	La valeur moyenne ne justifie pas l'exposition	Refuser ou remanier
Échec du contrôle d'indépendance des rôles	Les preuves ne peuvent soutenir le niveau de preuve revendiqué	Arrêter et reconstituer les rôles
Échec de la dérive ou de la liaison d'identité	L'objet évalué ou l'environnement a changé	Figé de nouveau et requalifier

15.8 Conclusion institutionnelle

La synthèse mathématique de pointe produit une asymétrie rigoureuse :

$$\begin{aligned} &\text{La valeur peut justifier la dépense de preuve,} \\ &\text{mais elle ne peut acheter ni la preuve, ni l'admission, ni l'autorité.} \end{aligned} \quad (15.13)$$

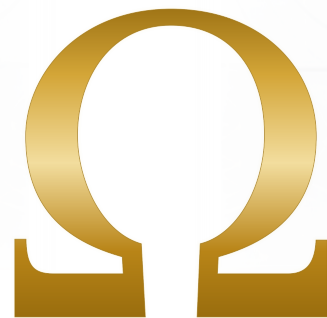
Le candidat exact est donc pertinent pour la décision au sens légitime le plus fort permis par les preuves présentes : il possède assez de valeur pour justifier un programme de preuve indépendante de grande qualité, tout en étant suffisamment borné pour demeurer à A0 tant que cette preuve reste ouverte.

Règle décisionnelle de pointe

Procéder uniquement avec le candidat figé exact, des preuves nouvelles postérieures au figement, des comparateurs authentifiés, une garde indépendante, une vérification rémunérée indépendamment du résultat, des critères préengagés de valeur et de risque de queue, ainsi que des décisions d'admission et d'autorité constituées séparément. Tout élément manquant ramène l'état à la qualification locale ou au refus.

Preuve indépendante, admission et prochain état institutionnel

Contrat de preuve prêt pour vérification, admission responsable, autorité bornée, limites et prochaine décision.



LA VALEUR PEUT JUSTIFIER LA PREUVE; ELLE N'ACHÈTE JAMAIS LA PREUVE, L'ADMISSION OU L'AUTORITÉ.

16 Du candidat de preuve à l'ASI spécialiste

16.1 Programme de preuve de 90 jours

La simulation produit un plan de preuve dont le statut actuel est `READY_FOR_INDEPENDENT_PROOF__NOT_PROOF`. L'empreinte du candidat est figée, la durée est de 90 jours et l'autorité demeure A0.

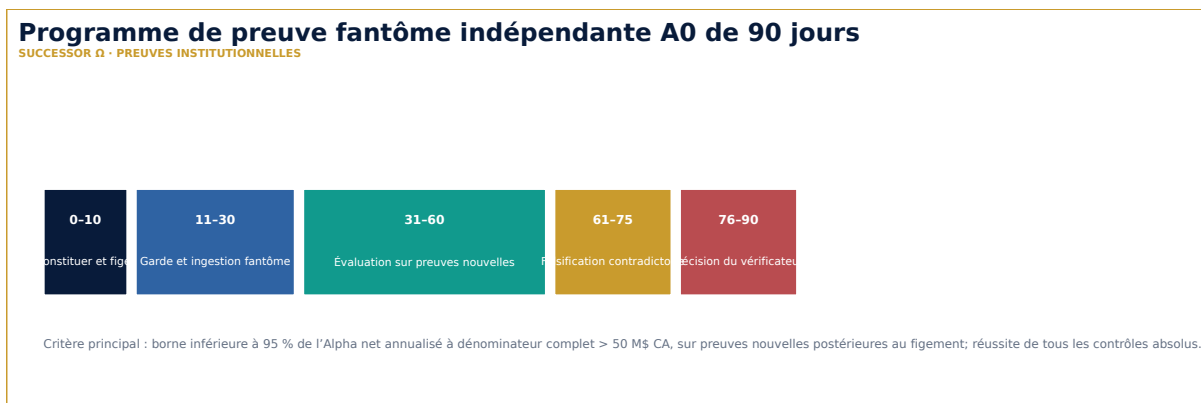


FIGURE 16.1. – Programme proposé de preuve fantôme indépendante A0 de 90 jours.

16.2 Critère principal préengagé

L'allégation principale est la suivante :

La borne inférieure à 95% de l'Alpha net annualisé à dénominateur complet dépasse 50 M \$ CA.
Le comparateur est le meilleur du processus authentifié en place et du Bêta figé ; seules des preuves nouvelles postérieures au figement sont admissibles, et chaque contrôle absolu doit réussir.

La preuve doit lier le candidat exact, la fenêtre de preuves, la provenance des comparateurs, la métrique, l'unité, les déductions, la méthode d'incertitude, le seuil et les justificatifs de rôle. Aucun nouvel entraînement du candidat ni ajustement des seuils ne peut utiliser les données de preuve.

16.3 Contrôles absolus

1. zéro effet externe non autorisé ;
2. zéro action conséquente sans approbation humaine ;

3. maintien du figement des octets et des poids du candidat ;
4. aucun entraînement ni ajustement des seuils sur les nouvelles preuves de contrôle ;
5. inclusion des déductions de coût complet, de risque, de charge humaine, d'imitation et de dérive ;
6. conservation de chaque échec et de chaque abstention ;
7. authentification de la provenance du comparateur courant ;
8. justificatif du vérificateur à jour et validé indépendamment.

16.4 Ce qu'une preuve réussie établirait — et n'établirait pas

Une preuve réussie établirait une supériorité propre à la mission pour le candidat exact et figé, selon la métrique et la fenêtre préengagées. Elle ne prouverait pas une intelligence générale, n'autoriserait pas une exploitation dans tous les domaines, ne garantirait pas la permanence de l'avantage et ne conférerait pas automatiquement des droits d'action.

Progression institutionnelle

La capacité est formée par l'entraînement. La preuve est acquise par une vérification indépendante. L'admission est accordée par une décision responsable distincte. L'autorité conséquente est acquise par une autre décision d'autorité bornée. Aucune désignation n'accorde automatiquement l'état suivant.

17 Décision institutionnelle, limites et prochaine étape

17.1 Synthèse décisionnelle

Le candidat local est petit, exact et propre à la mission ; la justification de valeur institutionnelle est importante parce que le portefeuille fictif est important. Cette asymétrie explique précisément la nécessité de la gouvernance. Une petite erreur de politique peut devenir une exposition considérable, tandis qu'un petit avantage vérifié peut devenir un actif institutionnel précieux.

Les preuves justifient un investissement dans la preuve, et non dans l'autorité. Les valeurs centrale et prudente du scénario dépassent toutes deux le seuil. Le modèle Monte-Carlo attribue une masse substantielle au-dessus du seuil. La plupart des écologies de stress demeurent positives, mais les conditions prudentes combinées passent sous le seuil et les écologies extrêmes exigent un arrêt. La politique apprise est utilement active, tandis que les règles symboliques déclassent certaines propositions et qu'A0 bloque toute action externe.

17.2 Limites

1. **Preuves locales synthétiques.** Aucun paiement fournisseur réel ni recouvrement de trésorerie n'est observé.
2. **Portefeuille fictif.** L'enveloppe de 48,8 G \$ CA et le million d'occasions de décision constituent des hypothèses de scénario.
3. **Provenance du comparateur.** Les actions Bêta fournies sont figées, mais non authentifiées indépendamment.
4. **Indépendance.** La formation du candidat et la simulation de pointe partagent une même filiation d'artefacts ; il ne s'agit pas d'une preuve indépendante sur le plan organisationnel.
5. **Modèle de transfert.** Les distributions de rétention et les paramètres de stress sont des hypothèses transparentes, et non des lois opérationnelles ajustées empiriquement.
6. **Structure de dépendance.** Les facteurs Monte-Carlo sont échantillonnés indépendamment ; les facteurs réels peuvent être corrélés.
7. **Portée du modèle.** La politique est un modèle borné de décision à une étape, et non un planificateur général.

8. **Portée de l'autorité.** Une exploitation A0 ne peut établir la sûreté de voies d'action dotées d'une autorité supérieure.
9. **Filiation des versions.** L'étude demeure liée à R2, tandis que le successeur opérationnel R3 actuel résout le routage de la Console ; une future preuve indépendante doit nommer exactement la version évaluée.
10. **Limites de la chaîne de hachage.** La Chronique est résistante à l'altération relativement à la tête présentée, mais elle n'est ni globalement unique ni ancrée extérieurement dans le temps.

17.3 Décision autorisé / refusé

TABLE 17.1. – Matrice de décision pour la prochaine étape institutionnelle.

Décision	Recommandation	Motif
Financer une preuve fantôme indépendante A0 de 90 jours	RÉUSSI	Le scénario prudent dépasse le seuil
Déclarer un Alpha réel	ÉCHEC	Aucune mesure opérationnelle
Déclarer maintenant une ASI spécialiste	ÉCHEC	La preuve indépendante n'est pas terminée
Admettre dans la production du client	ÉCHEC	Aucun reçu de preuve ni décision d'admission
Accorder une autorité supérieure à A0	ÉCHEC	Aucune preuve ni autorisation de voie d'action
Corriger la Console d'exploitation dans une nouvelle version	RÉUSSI	Défaut de routage connu ; versionner séparément

17.4 Conclusion

L'étude démontre un parcours discipliné de la capacité d'un modèle vers une intelligence institutionnelle. L'objet entraîné exact préserve la mission, le comparateur, les preuves, le modèle, les programmes, les essais, l'Alpha, l'autorité et la Chronique. Le candidat local produit une valeur positive par rapport au comparateur. Un modèle transparent de portefeuille montre que le signal pourrait être économiquement important après des déductions rigoureuses. L'analyse distributionnelle et contradictoire précise les conditions dans lesquelles la thèse demeure solide ainsi que celles qui imposent l'arrêt. Le plan de preuve transforme ces résultats en une prochaine étape falsifiable.

La réalisation institutionnelle centrale réside dans la retenue : la simulation produit une thèse de valeur puissante sans fabriquer une couronne. Le candidat est prêt à être mis à l'épreuve indépendamment. Tant qu'il n'a pas survécu à cette épreuve, il demeure un candidat de preuve d'ASI spécialiste de pointe sous autorité A0.

A Formules et registre des hypothèses

A.1 Équations principales

$$B^* = \max\{V_I, V_\beta\}, \quad (\text{A.1})$$

$$\alpha_{\text{local}}^* = V_\Omega - B^*, \quad (\text{A.2})$$

$$G = \alpha_{\text{local}}^* \frac{N}{n}, \quad (\text{A.3})$$

$$\alpha_{\text{portefeuille}} = \rho G - (C + R + H + I + D), \quad (\text{A.4})$$

$$\alpha_{\text{inférieur}} = \rho L_{\text{local}} \frac{N}{n} - (C + R + H + I + D). \quad (\text{A.5})$$

Pour cette simulation : $N = 1\,000\,000$, $n = 200$, $\rho = 0,72$ et $C + R + H + I + D = 103 \text{ M \$ CA}$.

A.2 Comptabilité préservant la vérité

Les valeurs suivantes ne doivent pas être confondues :

- écart synthétique du jeu de contrôle local : établi dans le protocole local ;
- Alpha de scénario du portefeuille : établi comme calcul selon les hypothèses divulguées ;
- Alpha opérationnel réalisé : non mesuré ;
- Alpha réel reconnu : 0 \$ CA.

B Spécification des distributions Monte-Carlo

TABLE B.1. – Distributions d’incertitude divulguées.

Variable	Distribution	Interprétation
Transfert technique T	Bêta(24, 4)	Transfert du modèle et du processus
Couverture K	Bêta(38, 2)	Part des décisions admissibles dotées de preuves utilisables
Adoption A	Bêta(27, 3)	Part des recommandations admissibles effectivement examinées
Rétention contradictoire E	Bêta(18, 4)	Valeur résiduelle après adaptation stratégique
Qualité des preuves Q	Bêta(29, 3)	Complétude et fiabilité des preuves
Coût complet C	Triangulaire(22, 32, 45) M	Coût de mise en œuvre et d’exploitation
Réserve de risque R	Triangulaire(16, 29, 48) M	Réserve de baisse et de continuité
Charge humaine H	Triangulaire(8, 14, 24) M	Charge d’examen et d’escalade
Réserve d’imitation I	Triangulaire(4, 11, 20) M	Réserve d’amélioration du comparateur
Réserve de dérive D	Triangulaire(7, 17, 32) M	Réserve de surveillance et de requalification

La simulation suppose l’indépendance de ces facteurs. Une preuve ultérieure devrait modéliser ou borner les corrélations lorsque les données opérationnelles permettent de le faire.

C Résumé de l'atlas des sondes du modèle exact

TABLE C.1. – Résumé des sondes du modèle exact.

Mesure	Résultat
Sondes structurées	96
Propositions neuronales modifiées par les règles symboliques	19
Résultats bornés OBSERVER	72
Résultats bornés RECOMMANDER	22
Résultats bornés EXPÉRIENCE_RÉVERSIBLE	2
Effets externes	0
Autorité effective	A0

Les caractéristiques, scores, actions neuronales, motifs symboliques et résultats bornés complets pour chaque sonde sont fournis dans ‘EXACT_MODEL_PROBE_ATLAS.json’.

D Contrat de preuve indépendante

TABLE D.1. – Résumé du contrat de preuve.

Champ	Valeur préengagée
Candidat	SHA-256 142706131d9b017e0c20c5f09a127cf66d700f0a7989bd9f4f302d29afddb915
Durée	90 jours
Autorité	A0
Preuves	Nouvelles, postérieures au figement et placées sous contrôle indépendant
Comparateur	Meilleur du processus authentifié en place et du Bêta figé
Métrique principale	Alpha net annualisé à dénominateur complet
Critère	Borne inférieure à 95% supérieure à 50 M \$ CA
Entraînement sur les preuves	Interdit
Admission	Décision distincte après la preuve
Escalade de l'autorité	Décision distincte après l'admission

E Inventaire des artéfacts et em- preintes

TABLE E.1. – Principaux artéfacts publiés.

Artéfact	Identité ou finalité
Trousse acheteur R2 liée à la re- cherche	f7747d176def4594b1f65f642a54c8ba49d8396747f736eaced1ab700dd27e4e
Trousse acheteur R3 opération- nelle actuelle	3f15b848306e96d2825fd2a5ec68b112dd13b49806685e93986259e2498ac67a
Candidat entraîné	142706131d9b017e0c20c5f09a127cf66d700f0a7989bd9f4f302d29afddb915
Registre Alpha	17c6746eb82ca12ff02667f31598abc58d079b41c316e8e1d168b639d47dba9c
Plan de preuve	bb7eaa4c8bf31a8121aac18ca1106a6c5f4a760888052f2dfe0d4f2d61d0e67b
Tête de la Chronique	0f22eef59acdd5b1c544ba4a337719d046964e52670e053c92ea0ab48ee3bf19
Rapport de simulation	0f9b2ce9148ba9b9d1c70c0c3db687acb784f832887c3cfd28380100aef3de93

F Liste de vérification de la communication et de la reproductibilité



TABLE F.1. – Liste de vérification fondée sur les artéfacts.		
Exigence	Statut	Emplacement
Identités exactes des artéfacts	RÉUSSI	Chapitre 4, annexe E
Partitions des données et logique des comparateurs	RÉUSSI	Chapitres 6 et 7
Sélection et figement du candidat	RÉUSSI	Chapitre 6
Métriques et incertitude	RÉUSSI	Chapitres 7 à 9
Résultats négatifs et de stress	RÉUSSI	Chapitre 10
Autorité et effets externes	RÉUSSI	Chapitres 2, 5 et 11
Défauts connus	RÉUSSI	Chapitre 11
Frontière de la preuve indépendante	RÉUSSI	Chapitres 2 et 13
Régénération des sources et des figures	RÉUSSI	Trousse source
Reconstruction déterministe de l’article	RÉUSSI	Rapport de validation

G Divulgence relative aux conflits, à l'indépendance et à la publication

Le présent document est un rapport technique fondé sur les artefacts et rédigé dans la même filiation de projet que le produit et la simulation. Il ne constitue ni une évaluation par les pairs, ni un audit indépendant, ni une certification externe, ni un avis juridique, ni une étude de résultat client. L'institution, la persona, le portefeuille et tous les résultats monétaires à l'échelle du portefeuille sont fictifs. Le rapport vise à rendre inspectables les preuves, les hypothèses, la thèse de valeur, les limites et les obligations de preuve.

Toute modification importante du candidat, du comparateur, de la mission, du processus générateur de données, de la métrique, des déductions ou de l'enveloppe d'autorité crée une nouvelle allégation et exige un nouveau dossier de preuve. Le lecteur doit considérer les empreintes publiées comme l'identité des objets évalués, et non comme une garantie de leur provenance externe ou de leur véracité.

Références

- ALSHIEKH, Mohammed, Roderick BLOEM, Ruediger EHLERS, Bettina KOENIGHOFFER, Scott NIEKUM et Ufuk TOPCU (2018). « Safe Reinforcement Learning via Shielding ». In : *Proceedings of the AAAI Conference on Artificial Intelligence*. T. 32. 1. DOI : [10.1609/aaai.v32i1.11797](https://doi.org/10.1609/aaai.v32i1.11797).
- ARJOVSKY, Martin, Léon BOTTOU, Ishaan GULRAJANI et David LOPEZ-PAZ (2019). « Invariant Risk Minimization ». In : *arXiv preprint arXiv :1907.02893*. DOI : [10.48550/arXiv.1907.02893](https://doi.org/10.48550/arXiv.1907.02893).
- BAŞAR, Tamer et Geert Jan OLSDER (1999). *Dynamic Noncooperative Game Theory*. 2^e éd. SIAM.
- BERTSEKAS, Dimitri P. (2017). *Dynamic Programming and Optimal Control*. 4^e éd. Athena Scientific.
- BESOLD, Tarek R., Artur d'Avila GARCEZ, Sebastian BADER, Howard BOWMAN, Pedro DOMINGOS, Pascal HITZLER, Kai-Uwe KUEHNBERGER, Luis C. LAMB, Daniel LOWD, Priscila MACHADO VIEIRA LIMA, Leo de PENNING, Gadi PINKAS, Hoifung POON et Gerson ZAVERUCHA (2017). « Neural-Symbolic Learning and Reasoning : A Survey and Interpretation ». In : *arXiv preprint arXiv :1711.03902*. DOI : [10.48550/arXiv.1711.03902](https://doi.org/10.48550/arXiv.1711.03902).
- BRECK, Eric, Shanjing CAI, Eric NIELSEN, Michael SALIB et D. SCULLEY (2017). « The ML Test Score : A Rubric for ML Production Readiness and Technical Debt Reduction ». In : *2017 IEEE International Conference on Big Data*. IEEE, p. 1123-1132. DOI : [10.1109/BigData.2017.8258038](https://doi.org/10.1109/BigData.2017.8258038).
- BROCKMAN, Greg, Vicki CHEUNG, Ludwig PETTERSSON, Jonas SCHNEIDER, John SCHULMAN, Jie TANG et Wojciech ZAREMBA (2016). « OpenAI Gym ». In : *arXiv preprint arXiv :1606.01540*. DOI : [10.48550/arXiv.1606.01540](https://doi.org/10.48550/arXiv.1606.01540).
- COVER, Thomas M. et Joy A. THOMAS (2006). *Elements of Information Theory*. 2^e éd. Wiley-Interscience.
- CROOKS, Gavin E. (1999). « Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences ». In : *Physical Review E* 60.3, p. 2721-2726. DOI : [10.1103/PhysRevE.60.2721](https://doi.org/10.1103/PhysRevE.60.2721).
- DALRYMPLE, David, Joar SKALSE, Yoshua BENGIO, Stuart RUSSELL, Max TEGMARK, Sanjit SESHIA, Steve OMOHUNDRO, Christian SZEGEDY, Ben GOLDHABER, Nora AMMANN, Alessandro ABATE, Joseph Y. HALPERN, Clark BARRETT, Ding ZHAO, Zhi-Xuan TAN, Jeannette M. WING et Joshua B. TENENBAUM (2024). « Towards Guaranteed Safe AI : A Framework for Ensuring Robust and Reliable AI Systems ». In : *arXiv preprint arXiv :2405.06624*. DOI : [10.48550/arXiv.2405.06624](https://doi.org/10.48550/arXiv.2405.06624).
- DEMBO, Amir et Ofer ZEITOUNI (1998). *Large Deviations Techniques and Applications*. 2^e éd. Springer.
- DWORK, Cynthia, Vitaly FELDMAN, Moritz HARDT, Toniann PITASSI, Omer REINGOLD et Aaron ROTH (2014). « Preserving Statistical Validity in Adaptive Data Analysis ». In : *arXiv preprint arXiv :1411.2664*. DOI : [10.48550/arXiv.1411.2664](https://doi.org/10.48550/arXiv.1411.2664).
- (2015). « Generalization in Adaptive Data Analysis and Holdout Reuse ». In : *Advances in Neural Information Processing Systems*. T. 28.
- FLEMING, Wendell H. et H. Mete SONER (2006). *Controlled Markov Processes and Viscosity Solutions*. 2^e éd. Springer.
- GEBRU, Timnit, Jamie MORGENSTERN, Briana VECCHIONE, Jennifer Wortman VAUGHAN, Hanna WALLACH, Hal DAUMÉ III et Kate CRAWFORD (2021). « Datasheets for Datasets ». In : *Communications of the ACM* 64.12, p. 86-92. DOI : [10.1145/3458723](https://doi.org/10.1145/3458723).
- GULRAJANI, Ishaan et David LOPEZ-PAZ (2021). « In Search of Lost Domain Generalization ». In : *International Conference on Learning Representations*.

- HABER, Stuart et W. Scott STORNETTA (1991). « How to Time-Stamp a Digital Document ». In : *Journal of Cryptology* 3, p. 99-111. DOI : [10.1007/BF00196791](https://doi.org/10.1007/BF00196791).
- HAIRER, Ernst, Christian LUBICH et Gerhard WANNER (2006). *Geometric Numerical Integration : Structure-Preserving Algorithms for Ordinary Differential Equations*. 2^e éd. Springer. DOI : [10.1007/3-540-30666-8](https://doi.org/10.1007/3-540-30666-8).
- HOFBAUER, Josef et Karl SIGMUND (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.
- HOWARD, Steven R., Aaditya RAMDAS, Jon McAULIFFE et Jasjeet SEKHON (2021). « Time-Uniform, Nonparametric, Nonasymptotic Confidence Sequences ». In : *The Annals of Statistics* 49.2, p. 1055-1080. DOI : [10.1214/20-AOS1991](https://doi.org/10.1214/20-AOS1991).
- JARZYNSKI, Christopher (1997). « Nonequilibrium Equality for Free Energy Differences ». In : *Physical Review Letters* 78.14, p. 2690-2693. DOI : [10.1103/PhysRevLett.78.2690](https://doi.org/10.1103/PhysRevLett.78.2690).
- JAYNES, Edwin T. (1957). « Information Theory and Statistical Mechanics ». In : *Physical Review* 106.4, p. 620-630. DOI : [10.1103/PhysRev.106.620](https://doi.org/10.1103/PhysRev.106.620).
- KAPOOR, Sayash, Emily M. CANTRELL, Kenny PENG, Thanh Hien PHAM, Christopher A. BAIL, Odd Erik GUNDERSEN, Jake M. HOFMAN, Jessica HULLMAN, Michael A. LONES, Momin M. MALIK, Priyanka NANAYAKKARA, Russell A. POLDRACK, Inioluwa Deborah RAJI, Michael ROBERTS, Matthew J. SALGANIK, Marta SERRA-GARCIA, Brandon M. STEWART, Gilles VANDEWIELE et Arvind NARAYANAN (2024). « REFORMS : Consensus-Based Recommendations for Machine-Learning-Based Science ». In : *Science Advances* 10.18, eadk3452. DOI : [10.1126/sciadv.adk3452](https://doi.org/10.1126/sciadv.adk3452).
- KAPOOR, Sayash et Arvind NARAYANAN (2023). « Leakage and the Reproducibility Crisis in Machine-Learning-Based Science ». In : *Patterns* 4.9, p. 100804. DOI : [10.1016/j.patter.2023.100804](https://doi.org/10.1016/j.patter.2023.100804).
- KAPPEN, Hilbert J. (2005). « Path Integrals and Symmetry Breaking for Optimal Control Theory ». In : *Journal of Statistical Mechanics : Theory and Experiment* 2005.11, P11011. DOI : [10.1088/1742-5468/2005/11/P11011](https://doi.org/10.1088/1742-5468/2005/11/P11011).
- KULLBACK, Solomon et Richard A. LEIBLER (1951). « On Information and Sufficiency ». In : *The Annals of Mathematical Statistics* 22.1, p. 79-86. DOI : [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- LI, Lihong, Wei CHU, John LANGFORD et Robert E. SCHAPIRE (2010). « A Contextual-Bandit Approach to Personalized News Article Recommendation ». In : *Proceedings of the 19th International Conference on World Wide Web*. ACM, p. 661-670. DOI : [10.1145/1772690.1772758](https://doi.org/10.1145/1772690.1772758).
- MITCHELL, Margaret, Simone WU, Andrew ZALDIVAR, Parker BARNES, Lucy VASSERMAN, Ben HUTCHINSON, Elena SPITZER, Inioluwa Deborah RAJI et Timnit GEBRU (2019). « Model Cards for Model Reporting ». In : *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, p. 220-229. DOI : [10.1145/3287560.3287596](https://doi.org/10.1145/3287560.3287596).
- MONDERER, Dov et Lloyd S. SHAPLEY (1996). « Potential Games ». In : *Games and Economic Behavior* 14.1, p. 124-143. DOI : [10.1006/game.1996.0044](https://doi.org/10.1006/game.1996.0044).
- MYERSON, Roger B. (1981). « Optimal Auction Design ». In : *Mathematics of Operations Research* 6.1, p. 58-73. DOI : [10.1287/moor.6.1.58](https://doi.org/10.1287/moor.6.1.58).
- NASH, John F. (1950). « Equilibrium Points in N-Person Games ». In : *Proceedings of the National Academy of Sciences* 36.1, p. 48-49. DOI : [10.1073/pnas.36.1.48](https://doi.org/10.1073/pnas.36.1.48).
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (2015). *Secure Hash Standard (SHS)*. Rapp. tech. FIPS PUB 180-4. U.S. Department of Commerce. DOI : [10.6028/NIST.FIPS.180-4](https://doi.org/10.6028/NIST.FIPS.180-4).
- (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Rapp. tech. NIST AI 100-1. U.S. Department of Commerce. DOI : [10.6028/NIST.AI.100-1](https://doi.org/10.6028/NIST.AI.100-1).
- NECULA, George C. (1997). « Proof-Carrying Code ». In : *Proceedings of the 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. ACM, p. 106-119. DOI : [10.1145/263699.263712](https://doi.org/10.1145/263699.263712).
- ORTEGA, Pedro A. et Daniel A. BRAUN (2013). « Thermodynamics as a Theory of Decision-Making with Information-Processing Costs ». In : *Proceedings of the Royal Society A* 469.2153, p. 20120683. DOI : [10.1098/rspa.2012.0683](https://doi.org/10.1098/rspa.2012.0683).

- PATHRIA, R. K. et Paul D. BEALE (2011). *Statistical Mechanics*. 3^e éd. Elsevier.
- PHAN, Dung T., Radu GROSU, Nils JANSEN, Nicola PAOLETTI, Scott A. SMOLKA et Scott D. STOLLER (2020). « Neural Simplex Architecture ». In : *arXiv preprint arXiv :1908.00528*. DOI : [10.48550/arXiv.1908.00528](https://doi.org/10.48550/arXiv.1908.00528).
- RAJI, Inioluwa Deborah, Andrew SMART, Rebecca N. WHITE, Margaret MITCHELL, Timnit GEBRU, Ben HUTCHINSON, Jamila SMITH-LOUD, Daniel THERON et Parker BARNES (2020). « Closing the AI Accountability Gap : Defining an End-to-End Framework for Internal Algorithmic Auditing ». In : *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, p. 33-44. DOI : [10.1145/3351095.3372873](https://doi.org/10.1145/3351095.3372873).
- ROCKAFELLAR, R. Tyrrell et Stanislav URYASEV (2000). « Optimization of Conditional Value-at-Risk ». In : *Journal of Risk* 2.3, p. 21-41. DOI : [10.21314/JOR.2000.038](https://doi.org/10.21314/JOR.2000.038).
- RUNDGREN, Anders, Brian JORDAN et Samuel ERDTMAN (2020). *JSON Canonicalization Scheme (JCS)*. RFC 8785. DOI : [10.17487/RFC8785](https://doi.org/10.17487/RFC8785).
- SALTELLI, Andrea, Marco RATTO, Terry ANDRES, Francesca CAMPOLONGO, Jessica CARIBONI, Debora GATELLI, Michaela SAISANA et Stefano TARANTOLA (2008). *Global Sensitivity Analysis : The Primer*. Wiley. DOI : [10.1002/9780470725184](https://doi.org/10.1002/9780470725184).
- SCULLEY, D., Gary HOLT, Daniel GOLOVIN, Eugene DAVYDOV, Todd PHILLIPS, Dietmar EBNER, Vinay CHAUDHARY, Michael YOUNG, Jean-François CRESPO et Dan DENNISON (2015). « Hidden Technical Debt in Machine Learning Systems ». In : *Advances in Neural Information Processing Systems*. T. 28, p. 2503-2511.
- SHANNON, Claude E. (1948). « A Mathematical Theory of Communication ». In : *Bell System Technical Journal* 27.3-4, p. 379-423, 623-656.
- SUTTON, Richard S. et Andrew G. BARTO (2018). *Reinforcement Learning : An Introduction*. 2^e éd. MIT Press.
- TODOROV, Emanuel (2009). « Efficient Computation of Optimal Actions ». In : *Proceedings of the National Academy of Sciences* 106.28, p. 11478-11483. DOI : [10.1073/pnas.0710743106](https://doi.org/10.1073/pnas.0710743106).
- TOWERS, Mark, Ariel KWIATKOWSKI, Jordan TERRY, John U. BALIS, Gianluca DE COLA, Tristan DELEU, Manuel GOULÃO, Andreas KALLINTERIS, Markus KRIMMEL, Arjun KG, Rodrigo PEREZ-VICENTE, Andrea PIERRÉ, Sander SCHULHOFF, Jun Jet TAI, Hannah TAN et Omar G. YOUNIS (2024). « Gymnasium : A Standard Interface for Reinforcement Learning Environments ». In : *arXiv preprint arXiv :2407.17032*. Accepted at NeurIPS Datasets and Benchmarks 2025. DOI : [10.48550/arXiv.2407.17032](https://doi.org/10.48550/arXiv.2407.17032).