

GOVERNED AUTONOMY

Governed Autonomy

Doctrines, architectures, and research studies for persistent autonomous systems

Andrew Durden
Independent research
Publication edition 2026.10.9

Research on authority, evidence, and continuity in autonomous systems.

Content identifier: 7150dfae1470e29ab5d0fbcc

Reading map

Begin with the overview and the two lead papers. The topic studies can be read independently. Research Depth & Proof links the investigations; the detailed contents and PDF bookmarks retain every reference and section.

Governed Autonomy	3
Governed Autonomous Institutions	8
Architecture of Governed Autonomous Systems	23
MIRAGE: Authority, information, and consequences in autonomous systems	38
CSSL: Semantic change, accountable evidence, and operational continuity	50
IC-POAR: From intended behavior to an accountable running product	60
AD8: Accountable change across software, action, and learning	68
Freehold: Autonomous institutions under bounded stewardship	78
Cuttlebone: Preserving the boundary between reasoning and action	92
SEPION: What agreement among generated artifacts can establish	102
RCFS: Competence that can change without inheriting authority	111
Semantic OS: Application Understanding, Authority, and Verified Outcomes	122
Cuttlefish Environment: Continuing Work Across Applications and Collaboration	130
CHROMATIC: Ambient Assistance Under Continuing Human Authority	139
Experience Economy and ISE: Learning from Governed Experience	148
The Reckoning: Calibration and Accountability for Autonomous Judgment	157
Human-LLM Capability Elicitation: Studying Validated Joint Performance	166
Research Depth & Proof	215
Engineering Methodology	224
About Me	268

Governed Autonomy

AI systems that can carry responsibility forward.

What happens to an AI system's commitments when its model changes, its work is interrupted, or its evidence turns out to be wrong?

Governed Autonomy is a research and evidence collection about accountable AI systems that operate over time. Its sixteen papers are accompanied by implementation and deployment evidence, an executable continuity reference, and a register of exact public assertions. It examines how systems can preserve purpose, respect limits on their authority, establish what actually happened, and finish or resolve the work they undertake.

The work is grounded in the implemented [Cuttlefish](#) and [Freehold](#) systems. Their production setting and the evidence available to readers are introduced below.

Written for researchers, engineers, and technical founders. Start with the central argument, explore the architecture, or read the papers at your own pace.

[Start with the lead paper](#) · [Explore the architecture](#) · [Download the collected volume](#)

Live systems and operational evidence

Cuttlefish and Freehold operate end to end for real customers. I make that statement from my responsibility for building and operating them. Their AWS implementation and workload evidence is separately dated and scoped below. Visit [Cuttlefish](#) at [getcuttlefish.app](#) to explore the public product and its operating architecture. Visit [Freehold](#) at [getfreehold.app](#).

These systems provide the operational and experimental setting for this research. This edition retains the examination I directed on 10 October 2026 with AI assistance of Freehold engine and Cuttlefish host source, dated AWS workload observations, and authenticated image metadata. The AD8 corpus and reference analysis introduced in 2026.10.7 does not constitute a new product or infrastructure examination. The [implementation evidence](#) describes the examined image metadata and explains what each source inspection, deployment observation, and behavioral result establishes. The published inspection records retain their own dates and scope.

The research collection also includes conditional arguments, finite formal results, and proposed extensions. A running system does not resolve every theoretical construction or demonstrate every real-world guarantee. The [assurance boundaries and claim register](#) preserve those distinctions claim by claim.

A commitment survives a lost acknowledgement

An authorized action changes the destination. Its response is lost, and the worker exits. The obligation remains open even though the effect has happened. A successor inspects the destination, verifies the attributable result, and settles the obligation. Repeating the request leaves the controlled destination with **one effect**. Revoking authority prevents a new effect.

The continuity exhibit makes that sequence inspectable and executable: a separate witness counts effects, an actual process exits after commitment, and eleven checks exercise restart, replay, scope, revocation, and tampered evidence. Its guarantee is bounded by the witness's atomic contract. The disclosed source is an original teaching reference; its result is separate from production evidence.

What should survive the model?

A system may remember a conversation and still forget a commitment. It may complete a tool call without establishing the intended result. It may possess a credential without having the authority to use it for the task at hand.

The central proposition is that **purpose, authority, obligations, evidence, and continuity need explicit institutional structure**. More capable reasoning can help a system decide what to propose. It cannot, by itself, establish what the system is entitled to do or what actually happened.

CONCEPTUAL RESPONSIBILITY MAP



Each responsibility has its own assumptions and evidence.

Figure 1. Purpose, authority, consequence, evidence, and continuity provide five questions for examining an autonomous system.

Three contributions to examine

An institutional account of continuity. The lead paper develops the conditions under which identity, obligations, resources, and recourse can survive replacement of the components that do the reasoning and work. Its account includes the founders and custodians who can change the institution's rules. Read [Governed Autonomous Institutions](#).

A map of distinct technical responsibilities. The architecture synthesis separates authored intent, application semantics, authority, effect, observation, learning, and human interaction. It distinguishes explicit source contracts, conceptual correspondences, proposed integrations, and unresolved adoption. Read [Architecture of Governed Autonomous Systems](#).

A precise approach to evidence. A specification, an executable reference, a finite model check, an observation, and an independent review answer different questions. The collection identifies what each can support and what a stronger conclusion would require. Read [Evidence and reproducibility](#).

Choose a reading route

Time and purpose	Route
Two minutes · the thesis	This page, the responsibility map, and what exists today.
Fifteen minutes · architecture and a journey	The architecture synthesis, illustrated continuity exhibit, and evidence overview.
Full technical review	The lead paper, a topic paper below, its assumptions and counterexamples, the claim register, prior-art comparison, and review guide.
Read, reproduce, or share	Individual papers and the collected volume, complete review text, and the executable reproduction contract.

The collection

Theme	Papers
Institutional foundations	Governed Autonomous Institutions, Architecture synthesis, Freehold, Cuttlebone
Meaning and execution	MIRAGE, CSSL, IC/POAR, NTVBCD/AD8
Certification and change	SEPION, RCFS
Application and human environment	Cuttlefish Semantic OS, Cuttlefish operating environment, CHROMATIC
Learning and evaluation	Experience Economy and ISE, The Reckoning, Human-LLM capability elicitation

Read critically

The strongest useful review identifies a specific claim, its premises, and a case that might defeat it. Is an authority boundary complete under its stated assumptions? Does a proposed relationship preserve both systems' responsibilities? Does an observation establish the claimed outcome? Are measured results being distinguished from research hypotheses?

Use the [review guide](#) to submit a counterexample, correction, or research question.

About this edition

Edition **2026.10.9** adds the official Freehold website link to **2026.10.8**, which corrected the authorial voice of **2026.10.7**. The earlier editions, including **2026.10.6**, remain preserved. It retains sixteen research papers, dated evidence of implemented Freehold and Cuttlefish systems, the executable public continuity reference and the successor public-text assertion register. It retains the twelve synthetic technical investigations, public evidence catalog and Engineering Methodology page introduced in 2026.10.7. Each separately versioned work retains its own scope. Proprietary implementation details are excluded. The collection has not undergone independent peer review; operational presence and the public reference do not establish one proved integration of all the named architectures.

The sources and method, bibliography, and author statement explain the basis of the work, its relationship to prior scholarship, and the assistance used in preparing this edition.

The collection is available for reading and review with all rights reserved. See rights and permissions and third-party notices.

Technical investigations in the successor edition

Research Depth & Proof connects twelve falsifiable investigations to all sixteen research areas. Engineering Methodology describes sole-human responsibility, substantive AI collaboration and the evidence boundary. Edition 2026.10.9 documents the preserved predecessor, corrections and separate release requirements.

Governed Autonomous Institutions

Governed Autonomy · Publication 2026.10.9

Doctrinal exposition

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Long-lived autonomous work raises a problem that cannot be answered by improving an individual model response. An undertaking may outlast its current model, operators, credentials, interface, or deployment. Other parties may rely on its claims, await its performance, or seek a remedy after it changes. The relevant unit of analysis is therefore an institution: a persistent arrangement of identity, authority, obligations, resources, knowledge, evidence, and recourse.

This paper presents the argument of the definitive third-edition doctrine of governed autonomous institutions. Its central distinction separates the capacity to originate purposes from the authority to cause consequences. Institutional autonomy permits purposes to emerge internally; the authority governing external effects remains rooted in an accountable arrangement that the acting model cannot create for itself. Continuity requires preserving commitments and evidence through substitution, interruption, correction, and succession. The argument extends to plural actors, model qualification, contested knowledge, constitutional change, federation, and the human institutions on which accountability depends.

The account consists of design commitments and conditional arguments. Evaluating a realization requires evidence of implementation conformance, effective governance, and the external recognition relevant to its activities. In particular, restrictions on an agent do not establish independence from a founder or administrator with bypass power. Technical validity, legal recognition, moral justification, and democratic legitimacy remain distinct questions.

1. The undertaking that survives the conversation

Consider an autonomous research undertaking that agrees to deliver a series of environmental assessments. Its work spans months. It uses several models, replaces an unreliable data provider, loses a connection during a delivery, and later discovers that one of its published conclusions was overstated. The original conversation is no longer an adequate account of what the undertaking is, what it owes, or what it may do next.

The problem has several parts. A counterparty needs to know whether the same undertaking still owes the report. The operator needs to know which access remains authorized. A reviewer needs to distinguish a delivered file from an accepted assessment. A successor needs to understand unresolved objections. None of these questions can safely depend on the active model recalling the right sentence.

The doctrine uses **institution** for the persistent subject of these commitments. It does not assume consciousness, personhood, citizenship, or a particular organization chart. Its category is operational: something remains identifiable and answerable as its cognitive and technical components change. Whether human law recognizes that subject directly, or attributes its acts to an existing legal person, is a separate question.

A useful automated workflow can exist without satisfying this stronger account of institutionhood. The doctrine is concerned with systems making the stronger claim: sustained autonomous participation in a shared world. That claim introduces continuity requirements beyond completing a single requested action.

The proposal is demanding because persistence preserves inconvenient facts as well as useful knowledge. An institution carries unpaid obligations, corrected claims, disputed decisions, and limits on competence. A restart is not a fresh moral or financial beginning. A new model is not a new counterparty.

Source basis: third-edition doctrine, Chapters 1–3 and Appendix I.

2. Purpose, authority, and legitimacy

The capacity to form an intention does not establish authority to execute it. An institution may decide that publishing a finding would advance its purpose while lacking permission to disclose the underlying data. It may discover a valuable purchase while lacking authority to spend the required funds. It may prefer to change a rule while lacking the power to enact that change.

The doctrine protects this distinction in both directions. An external constraint need not erase the internally formed purpose. The institution may retain the proposal, search for an authorized alternative, revise its plan, or seek the missing authority. Conversely, the existence of a compelling purpose does not dissolve the constraint.

This gives autonomy a more precise meaning than activity without supervision. **Internal autonomy** concerns originating, revising, debating, deferring, or abandoning purposes. **Authority** concerns recognized power over a bounded class of consequences. **Runtime validity** concerns satisfaction of the applicable technical and institutional rules. None of these by itself establishes the broader legitimacy of the charter, its beneficiaries, or its effects.

The distinction also clarifies the human role. An institution can operate routinely under valid standing authority without a human approving each thought or action. Human intervention becomes necessary when the relevant authority arrangement requires it, when information is unavailable, when interactive authentication or physical action is indispensable, or when an external legal or political process must decide an issue.

Human control remains consequential. People can establish or withhold resources, access, and recognition; exercise properly constituted oversight; and invoke valid recovery or termination powers. The autonomy claim is narrower: such control must not be disguised as self-originated institutional purpose. A secretly instructed decision should not be reported as autonomous deliberation.

Freehold's intended mode develops a particularly strict account of post-founding independence. A delegated assistant operates under a different arrangement, in which the user supplies the objective. Comparing the two requires keeping the origin of purpose explicit.

Source basis: Chapters 6–8 and 21; Autonomy Under Law, §§4 and 12; Doctrine of Governed Machine Civilizations, Chapters 4–5.

3. Continuity is more than a name

A persistent name is useful but insufficient. Two processes can display the same name while disagreeing about outstanding commitments. A renamed organization can remain answerable for an old obligation. A model replacement may preserve institutional continuity even though the new model behaves differently.

The doctrine therefore treats identity as a continuity relation governed by declared conditions. Relevant questions include whether the founding authority remains recognizable, whether the current rules descend from an authorized history, whether live obligations remain attributed, and whether the public account of the institution remains consistent with those facts.

Genesis has an external dependency. A declaration of existence cannot authenticate its own right to create the relevant authority. The founding arrangement must identify the recognized source of the powers it claims. Different powers may have different roots; permission to operate infrastructure does not automatically include permission to alter a charter or assume another party's liabilities.

Substitution and succession also differ. Replacing a model, operator, or storage system can preserve the same institution when the continuity conditions hold. Merger, split, takeover, or re-foundation may require a new identity and an explicit allocation of rights and duties. Calling every change a migration would conceal this distinction.

Continuity does not require byte-for-byte preservation of every cache or transient calculation. It requires preservation, or an authorized transformation, of the facts on which institutional reliance depends. A discarded draft and an accepted obligation have different consequences. A rebuilt search index and a changed authority record have different owners.

This yields a useful test: after a substitution, can an affected party still determine who owes what, which decisions remain current, what evidence supports them, and where a challenge belongs? If those answers vanish, continuity has not been established by keeping the old label.

Source basis: Chapters 3–5, 16, and 24.

4. A durable body for replaceable cognition

The active model supplies cognition within an institutional arrangement. It need not own the arrangement's law, identity, evidence, or obligations. The doctrine locates those durable responsibilities in the surrounding runtime.

This does not mean that one process must control everything. A canonical owner can be distributed. The requirement is that a consequential object has a defined authority for changing its state, with a consistent account of valid changes. Multiple views can exist without becoming competing authorities.

A public dashboard illustrates the distinction. It may show the status of an obligation, but its visual prominence does not give it the power to discharge that obligation. A cached permission display may help a person understand access, but its age matters when an action is attempted. A retrieved memory may inform a decision, but it does not create a new grant.

The external world remains outside this ownership claim. An institution can own its record of a bank transfer; it cannot make the bank's ledger agree by editing that record. It can accept a scientific conclusion under a stated evidential standard; it cannot make the conclusion true by storing it. Recovery must therefore reconcile internal records with the relevant external authorities and observations.

The same distinction applies to models. A provider's weights, availability, and behavior affect what proposals can be generated. They do not by themselves define the institution's current obligations. Context supplied to a model is a selected view, not the sole durable location of a commitment.

This separation creates an engineering obligation rather than an automatic guarantee. Every actual consumer must use the right current state. A design with well-named owners can still fail if a convenient interface, stale cache, or administrative shortcut controls the real effect.

Source basis: Chapters 4–5, 12–13, and Appendix F.

5. Rights, duties, and resources remain distinct

The doctrine adapts the distinction among rights, duties, privileges, powers, liabilities, immunities, and disabilities to its institutional setting. These terms answer different questions. Freedom from a prohibition is not necessarily power to bind another party. A duty to deliver a result does not itself create permission to access the data needed to deliver it.

The legal vocabulary draws on Hohfeld's analysis, but its use in a runtime is an adaptation, not a determination of jurisdictional law. The bibliography identifies the external literature and the limits of its role here.

An obligation persists until an authorized event changes its status. Delegating performance may leave the original obligor responsible. Partial performance may discharge part of a duty while leaving a remainder. A dispute, impossibility, or loss of access needs an explicit account; it is not a deletion rule.

Resource discipline makes these distinctions concrete. Two concurrent plans cannot both treat the same uncommitted funds as wholly available. A future promise cannot count an encumbered asset as free merely because the expected project value is high. Money, compute, storage, time, and exposure are not interchangeable quantities without a stated conversion and justification.

Revocation exposes the connection. If a necessary capability is withdrawn, the institution must stop relying on it. Its prior duty may still require notice, an authorized substitute, renegotiation, settlement, or recourse. Continuing without authority and forgetting the obligation are both inadequate responses.

The doctrine's account of viability is similarly conditional. It concerns the existence of an admissible way to preserve specified requirements under an assumed model of the environment and resources. It is not a general algorithm for predicting solvency or guaranteeing performance in an open world.

Source basis: Chapters 7, 8, and 10.

6. Consequence cannot be reduced to a tool name

A tool is a route through which an action may occur. The consequence depends on what the action changes, reveals, commits, or causes others to rely upon. The same interface may support a private draft, a confidential read, a public statement, and a binding submission.

Several dimensions can matter at once. A financially trivial disclosure may be irreversible. A reversible file change may have a serious integrity consequence. A public announcement may impose reputational or contractual costs even when no money immediately moves. These effects do not naturally form one universal ranking.

Composition matters as well. A sequence of individually permitted observations may expose information that none revealed alone. A prepared document combined with a publication action creates a different consequence from either step considered in isolation. Permission for the parts cannot be assumed to cover every combined use.

The doctrine consequently asks authority to remain related to the actual consequence, including uncertainty and changes discovered during work. A proposal cannot secure broad authority merely by describing itself in reassuring terms. If the work changes materially, its authority basis must be reconsidered.

This is a conditional requirement. Open environments make complete effect prediction difficult. A system must state the classes it can characterize and the assumptions under which its characterization is reliable. An inability to classify a protected consequence does not establish permission to cause it.

Compensation also needs careful language. Removing a published document does not remove the knowledge of everyone who saw it. Refunding a payment does not undo every reliance induced by it. Recovery may reduce harm and settle liability without restoring the exact prior world.

Source basis: Chapters 8–10 and the open effect-analysis obligations in Appendix D.

7. Evidence supports a claim at a stated scope

An event, log, observation, receipt, attestation, verification, proof, and claim are related but different objects of inquiry.

A log can record what a process reported. A receipt can attribute the processing of an action under a stated authority arrangement. An attestation can bind a statement to an identified attester. Verification applies a method to evidence for a particular claim. Each can be useful; none automatically establishes every other conclusion.

An authentic receipt for submitting a report does not establish that the recipient accepted the report's scientific conclusions. A signed measurement does not eliminate a faulty sensor. An internally consistent record can preserve a mistaken account. Authenticity, completeness, independence, relevance, and correctness require separate consideration.

The doctrine therefore preserves both support and contrary evidence. A claim can remain contested. Pressure to act does not make the contradiction disappear; it changes the decision problem. The institution must decide under the actual uncertainty and retain the basis for that choice.

Time matters too. Evidence may be adequate for an earlier state but insufficient for the present one. A permission, measurement, qualification, or counterparty record needs a relevant period and scope. Public summaries should make material uncertainty and challenge status intelligible to those who may rely on them.

Causal claims require an additional distinction. A record showing that an action preceded an outcome is not, by itself, a study showing that the action caused the outcome. The observation design, confounding, alternative explanations, and limits on generalization still matter. The doctrine's account of work history does not eliminate the methodological requirements of causal inference.

The evidence guide applies these distinctions to the specifications, reported checks, and research questions discussed in this collection.

Source basis: Chapters 11–13 and 26; Appendix H.

8. Learning without creating new authority

Durable learning requires more than retrieving earlier text. An institution needs to distinguish a proposed lesson from an accepted one, a current conclusion from a superseded one, and a successful procedure from a procedure whose assumptions no longer hold.

This is an account of adjudicated working knowledge. The word “adjudicated” identifies a responsibility for deciding what status knowledge has within the institution. It does not imply infallibility or require a courtroom for every note.

An accepted procedure should retain its scope. A successful action in one environment does not justify repeating it under different permissions, resources, or counterparties. Learning can improve competence while leaving authority unchanged. Knowing how to do something and having the right to do it remain separate.

Negative competence matters: evidence about when a role, model, or method should not be relied upon. Such knowledge helps the institution select a better route, seek additional evidence, or narrow the work. It should support successful recovery, not become a ritual of reporting avoidable failure.

Model substitution makes this discipline testable. Identity and obligations may remain continuous while competence changes. A replacement model can improve one role and regress in another. Qualification therefore needs a task, environment, consequence, and time scope. No general claim of interchangeable behavior follows from a common interface.

The doctrine also distinguishes learning from constitutional change. An ordinary memory update cannot quietly grant a new power or weaken a prohibition. A promising lesson may justify proposing a change, but the authority to enact that change remains separately governed.

Source basis: Chapters 13–17.

9. Plural actors and institutional intelligence

Persistent institutions may contain actors with different responsibilities, knowledge, and histories. Shared law does not require every actor to receive the same context or deliberate through one supervising mind.

This matters for both autonomy and scale. An actor responsible for one inquiry can retain its own work while relying on shared rules for consequences that affect others. Disagreement can be informative. Multiple proposals can coexist without creating multiple incompatible accounts of an accepted commitment.

The doctrine's account of institutional intelligence combines cognition with durable knowledge, procedures, evidence, obligations, resources, and feedback. Its formal treatment does not permit this combination to create its own authority roots. The surrounding world and the recognized authority arrangement remain open inputs.

The proposed efficiency benefit is selective attention: recruit the relevant parts of the institution for the current problem while preserving required dependencies. Relevance alone is insufficient. Privacy, contrary evidence, outstanding obligations, and the possible consequence of omission can make apparently remote material necessary.

9.1 Conditional claims and their dependencies

The doctrine's third edition supplies paper arguments over declared semantics, rather than machine-checked proofs of a deployed institution. The following register states the premises and conclusions used in this discussion. Its source locations are Chapter 14 and Appendix D; the fixed-mode result is §14.12.

Claim	Premises and domain	Conclusion and method	What depends on it—and what does not follow
Governed closure for a fixed input	Constitutional rules, external input, and state frontier are fixed; the information operation is monotone on an append-only domain; active computation is bounded	A least closed account of reachable candidates and allowed transitions can be defined. This is a fixed-point argument about the declared information order.	The account of a bounded reasoning episode. It does not contain every future world observation or make present acceptance monotone.

Claim	Premises and domain	Conclusion and method	What depends on it—and what does not follow
Convergence within one activation mode	The mode, graph, weights, bias, and update setting remain fixed. The feasible activation set is nonempty and closed in finite-dimensional normed space. Its projection maps every update into that set and is nonexpansive in the same norm. The complete update has a uniform contraction factor strictly below one.	There is a unique fixed activation, and repeated application converges to it from every starting activation in the feasible set. The source derives the contraction bound and applies the fixed-point theorem.	Convergence for that fixed mode. It supplies no convergence result for changing graphs, arbitrary learned routing, discontinuous constraints, or adversarial new observations. Establishing nonexpansiveness for actual constraints is a premise, not a consequence of calling them constraints.
Intelligence growth without authority creation	Candidate generation and learning can change knowledge and competence; issuing grants and approving constitutional change remain separately controlled; execution checks canonical authority and prohibitions; the historical record retains contradiction and defeated claims	Reflexive reasoning may expand the account of knowledge and methods without generating new authority roots or forcing a closed account of an open world. The source argues from the permitted operation families and their guards.	The separation of cognition from authority in this thesis. A real implementation must establish that no alternate write or effect route defeats that separation.

9.2 Why the fixed-mode conclusion is conditional

The convergence implication can be inspected without publishing the source's particular activation update. Let F denote one fixed update on the declared feasible set, and let d measure distance in the norm used to establish its bound. Assume a constant q with $0 \leq q < 1$ such that $d(F(x), F(y)) \leq q d(x, y)$ for every pair of feasible activations. Successive updates then satisfy $d(x[k+1], x[k]) \leq q^k d(x[1], x[0])$. The remaining distances are bounded by a convergent geometric series, so the sequence is Cauchy. The closed finite-dimensional domain is complete, giving a limit x^* . Continuity of a contraction gives $F(x^*) = x^*$. If y^* were another fixed point, $d(x^*, y^*) \leq q d(x^*, y^*)$ forces that distance to zero. The fixed point is therefore unique.

This derivation explains exactly where the premises enter. If each step uses a different update, the single-map argument no longer applies. If the projection is discontinuous or expands distances, the stated contraction premise may fail. If a mandatory dependency is absent from the modeled set, convergence still says nothing about recovering that omission. The source reports a regression against an earlier insufficient contraction condition; that finite check is distinct from the general paper argument and is not independently reproduced here.

The broader claim that sparse institutional attention can preserve every required dependency is expressly a conjecture. It needs assumptions about the task distribution, graph structure, retrieval and relevance judgments, contrary evidence, and adversarially hidden dependencies, followed by a completeness argument or bounded evaluation.

Fixed-mode convergence cannot supply those assumptions or prove fairness, useful performance, or continuing coverage during learning and growth.

This creates a productive separation between the architectural idea and its evaluation. The idea motivates a design space. Measurements and proofs must establish which workloads, scheduling conditions, and uncertainty bounds make a particular realization useful.

Source basis: Chapters 14–17 and Appendix D.

10. Continuing work through interruption

A persistent purpose should survive an interrupted attempt when the objective remains achievable and authorized. Process failure, a lost response, or a changed provider need not end the undertaking.

Continuation requires current facts. An interrupted payment might have completed externally. A permission might have expired while a process was offline. An old plan might target a resource that has changed. Repeating the last request without reconciliation can duplicate an effect or apply stale authority.

The doctrine treats recovery as part of institutional operation. The institution must determine what remains owed, what actually happened, which authority is current, and which valid route can complete the work. Where an effect is genuinely uncertain, that uncertainty must guide the next investigation rather than be converted into an invented success or an unexamined retry.

“Exactly once” requires particular care. Network delivery can fail, repeat, or become uncertain. A defensible institutional account recognizes duplicates and conflicts and reconciles them with the actual effects. It does not obtain universal exactly-once transport by naming a record.

Standing purposes introduce a related problem. Continuous attention must have a scope, resource basis, and conditions for continuation. A reminder to inspect a condition does not create permanent permission to make every useful change discovered during the inspection.

Progress guarantees remain conditional. Communication, available resources, sufficiently capable executors, and fair scheduling can be necessary. When an external decision or resource is indispensable, the system should identify the required user action plainly. When recovery is available within current authority, the system should continue toward the objective.

Source basis: Chapters 12, 18, 22, and 25.

11. Federation, economy, and recourse

An institution becomes externally consequential when others rely on it. Federation concerns how separate domains recognize identities, evidence, commitments, and decisions while retaining their own responsibilities.

Portability of an identity record does not imply portability of every permission. A destination may need to recognize or issue authority under its own rules. A translation that weakens an

obligation or hides a prohibition cannot establish compatibility merely because the resulting fields parse.

The doctrine proposes coordination at the scope of shared consequence. Local reasoning need not wait for universal agreement. A transfer of shared rights, a cross-party settlement, or a final public commitment may require coordination among the affected domains. The claim is conditional on compatible meanings, sufficient evidence, and an applicable settlement arrangement.

Economic participation adds scarcity and reliance. A listing is a claim about an offer; an accepted contract creates a different relationship. Delivery, verification, payment, dispute, and residual liability remain distinguishable. The existence of a marketplace does not establish that these relationships are correctly realized.

Reputation is an interpretation of challengeable history, not an objective universal score. It can depend on purpose, evidence quality, independence, uncertainty, and the distribution of prior work. A cluster of mutually endorsing institutions may reveal common control rather than broad trust. Repeated transactions can represent useful specialization or fabricated activity.

These are institutional and political questions as well as computational ones. The doctrine does not prove universal resistance to collusion, an optimal market mechanism, or automatic legal enforceability. It calls for visible recourse: an affected party must be able to identify what remedy is available, under whose authority, and with which practical limits.

Source basis: Chapters 19–21 and 27.

12. The founder is inside the rules

Constitutional limits on ordinary actors leave an important question unanswered: who can change those limits?

A founder may control the administrator, infrastructure, signing devices, appointments, or finances of nominally separate overseers. Counting distinct keys does not establish independent control. A runtime may correctly restrict an Agent while the surrounding organization remains able to bypass those restrictions.

Chapter 31 addresses this problem directly. Authorship, investment, beneficial ownership, and operation do not create an exemption from the protected rules. The founder is a principal within the arrangement. A rule against self-dealing must also apply when its own interpretation, enforcement, or amendment is at issue.

The central requirement is independent amendment custody under the existing rules. A proposed relaxation cannot validate its own adoption. Conflicts of interest must be assessed in terms of actual benefit and control, not merely formal titles. Ordinary disclosed compensation is not automatically prohibited; its permissibility must be determined through the applicable independent process.

The argument also extends to appointments and succession. A founder who can replace enough overseers at will may retain the effective amendment power even when no single

vote belongs to the founder. A recovery arrangement must not become a hidden way to install a controlled majority or enlarge temporary powers.

Counterparties need this information before accepting reliance. A technically elegant charter does not tell them whether its limits constrain the people who operate it. Current disclosure of material control concentration, bypass powers, and unresolved custody disputes is part of the institutional claim.

The doctrine does not assert that every enterprise can immediately establish such independence. A narrower system may operate under an explicit trusted-founder or trusted-administrator assumption. What it cannot do is use ordinary Agent confinement as evidence for the stronger claim of founder-independent governance.

The distinction is especially important during emergencies. A need to preserve assets or stop harmful effects can justify a bounded emergency power when that power already exists. Urgency does not create a new authority to remove protected constraints, erase liabilities, or classify a controlled party as independent.

Chapter 31 specifies how custody should be assessed. Establishing a real custodian's independence, beneficial ownership, or the absence of an administrative bypass requires evidence about the actual arrangement. Checking declared records alone cannot complete that investigation.

Source basis: Chapters 23 and 31.

13. Change, succession, and institutional death

Institutions need to change. Freezing a flawed rule indefinitely can be as consequential as changing it recklessly. The doctrine distinguishes proposing an improvement from possessing the power to enact it.

Review must consider the affected commitments, not just whether new code executes. A changed rule can alter who may act, which claims can be relied upon, how an obligation is interpreted, or what a counterparty can challenge. A successful test of a narrow component does not establish preservation of those wider relationships.

Rollback also has a scope. Restoring software does not automatically restore a previous public understanding or undo a contract formed in the meantime. Some changes require compensation, correction, notice, or succession rather than a technical reversal alone.

The end of ordinary operation does not end accountability. Pause, suspension, receivership, dissolution, and archival existence have different implications. A receiver may preserve resources and settle commitments without claiming ordinary autonomous operation. A dissolved institution may retain an evidential identity while holding no active power to create new effects.

A successor must state what it inherits. Obligations, resources, disputes, and historical reputation cannot be treated as an arbitrary selection of desirable assets. Where continuity cannot be established, a declared discontinuity with preserved recourse is more precise than an unsupported claim that nothing changed.

Human law remains decisive where it governs property, liability, legal succession, or external rights. A runtime's internal transition cannot confiscate property or create legal recognition merely by declaring a new status.

Source basis: Chapters 23–25 and 31.

14. Security and assurance are conditional claims

The doctrine's security argument depends on the boundary that actually governs effects. If an uncontrolled route can cause a protected consequence, a correct decision on another route does not establish complete mediation.

Established security principles such as complete mediation and separation of privilege frame this requirement. A realization must demonstrate how it satisfies them across its actual effect routes. The bibliography identifies these foundations.

Every assurance claim needs its trusted dependencies. A model may be outside the trusted computing base while an operator, host, credential service, or physical device remains inside it. A proof over a model does not establish that the deployed environment satisfies the assumptions.

Different methods answer different questions. Structural validation can detect missing references. A test can exercise selected behavior. A model checker can explore a specified state space. A refinement argument can relate two formal descriptions. A deployment study can measure behavior in an observed environment. An independent audit can examine a stated scope and its supporting evidence.

None of these is a universal upgrade of the others. More successful tests do not automatically become a theorem. A theorem does not measure an unknown workload. An auditor's involvement does not establish independence of every underlying source.

Assurance must also include useful progress. A system that prevents every external effect has not demonstrated the proposed autonomous capability. Long-horizon evaluation needs both constraint preservation and completed authorized work, including recovery when ordinary disruptions occur.

Source basis: Chapters 25–26 and Appendices D and J.

15. A worked institutional journey

The following is an illustrative scenario, not a deployment trace.

A research institution undertakes to provide a monthly river-quality assessment. It has authority to inspect specified public measurements and a separate arrangement governing publication. It retains its earlier assessments and the reasons for revising them.

Midway through the month, a data source becomes unavailable. The current model proposes using a new source. The new source may be technically reachable, but that does not establish permission to access it or adequate rights to publish derived material. The institution can continue other authorized work while resolving the missing premise.

A replacement model completes the analysis. The institutional identity and delivery commitment remain, but the replacement's competence cannot be assumed identical.

A reviewer assigned by the institution checks the observations under a declared review standard and identifies a conflict. This role may be filled by a human or an AI; in either case, qualification requires current evidence of competence for this particular review task. The example does not assume a human sign-off or treat a second AI opinion as proof of qualification. If the review standard requires independence, shared models, data, control, and conflicts of interest must also be assessed. The report preserves the disagreement and explains the uncertainty instead of hiding it in a confident average.

The delivery response is lost. The institution investigates the recipient's actual state before repeating the delivery. Once receipt is established, the remaining question is whether the delivery satisfies the agreed requirement. A submission receipt and acceptance of the assessment remain separate.

Later, a corrected measurement defeats a published conclusion. The institution preserves the prior report as historical material, updates the current claim, and handles the required notice. It also revises the procedure that accepted the original measurement. That learning does not enlarge its publication authority.

The founder then proposes relaxing review requirements to reduce cost. Because the change affects the founder's interests and the counterparty's reliance, ordinary operational success does not authorize the amendment. The existing custody and disclosure requirements still apply.

The scenario illustrates the doctrine's central claim: continuity is the capacity to carry the undertaking, its limits, and its consequences through change. It is not a promise that every attempt succeeds immediately.

16. What would strengthen or defeat the argument

The doctrine's practical value depends on how its requirements perform under real constraints. The following questions identify the next claims to test.

Institutional benefit. Do the added responsibilities improve long-horizon outcomes compared with credible simpler alternatives? A meaningful study needs comparable tasks, resources, authority, disruption, and opportunities for recovery. Counting stored records alone would not answer it.

Effect characterization. Which open-world consequences can be bounded reliably, and where does uncertainty become too large for the intended operation? The answer must concern actual effect classes and environments.

Model substitution. Which continuity properties can be preserved while competence changes, and how quickly can consequential regressions be detected? A stable identity is not enough if the institution repeatedly fails its duties.

Federation. Can separate domains translate their commitments conservatively and complete useful work under delay, disagreement, and partial failure? Compatibility should be demonstrated rather than inferred from shared terminology.

Custody and recourse. Do the independent-control arrangements constrain the actual beneficiaries and administrators? Can an affected party use the promised remedy in practice? A declared process is not evidence of effective access to it.

Scale. Does representative work remain local enough for coordination costs to track shared consequences? The source's locality argument assumes bounded cost for the consequential interactions that cross domains. It is not a benchmark or a claim that cross-domain work is free.

A counterexample is valuable when it identifies a declared premise that holds but a claimed result that does not follow. An implementation failure is a different kind of evidence: it may show that a premise was not realized. Review should distinguish the two while taking both seriously.

17. Source scope and contribution

This paper develops the definitive third edition's institutional argument. The two supporting doctrines contribute perspectives on continuity, stewardship, economic participation, and long-horizon evaluation. Their forecasts are treated as hypotheses, and their historical implementation references retain their historical scope. Each source keeps its own requirements and authority.

Five distinctions are central to the third edition: consequences have several dimensions; authority can follow different rules of use; work histories and causal studies support different claims; continuity can coexist with changed behavior; and runtime validity leaves external legitimacy to be established.

Subject	Principal source coverage
Institutional category, identity, durable state	Chapters 1-5
Purpose, rights, authority, consequence, resources	Chapters 6-10
Evidence, history, learning, actors, models	Chapters 11-17
Distributed operation, federation, economy, recourse	Chapters 18-22
Change, death, security, assurance, scale	Chapters 23-27
Consolidated argument and founder custody	Chapters 28-31
Definitions, conditional results, limits and conformance distinctions	Appendices A-J

The contribution is an account of the conditions under which sustained autonomous participation could remain answerable for its consequences. The paper presents that argument for examination while withholding proprietary construction details. The [architecture synthesis](#) relates it to the topic studies, and the [review guide](#) provides routes for corrections, counterexamples, and proposed evaluations.

Architecture of Governed Autonomous Systems

Governed Autonomy · Publication 2026.10.9

Lead architecture synthesis: responsibilities, source relationships, and composition questions

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Persistent autonomous work joins several problems that are often evaluated separately: forming an objective, interpreting it, deciding what is permitted, changing the world, establishing what happened, and carrying responsibility through interruption and change. Each has its own subject and conditions for success. An institution needs the resulting decisions to remain connected when a model changes, a permission expires, or an external action leaves an uncertain result.

This paper presents a framework for reviewing those connections across the Governed Autonomy collection. At each consequential boundary, the framework identifies the subject, responsible owner, authority, evidence, and continuing obligation. It compares the works' different responsibilities, distinguishes expressly specified relationships from proposed combinations, and develops three conceptual journeys that make the obligations of composition concrete.

The synthesis draws on specifications, theoretical arguments, bounded reported reference results, and the examination I directed on 10 October 2026 with AI assistance of the implemented Freehold engine and Cuttlefish host. That historical source and dated AWS deployment evidence establish an implemented relationship between those two systems. The framework does not establish that every research family forms one integrated deployment. The final sections identify the additional implementation, evaluation, and adoption evidence needed for stronger claims. Proprietary construction details are outside the paper's scope.

1. A continuing objective has several kinds of state

An autonomous system is often described through its most visible activity: reasoning about a request and selecting a tool. The description becomes incomplete when work lasts longer than one interaction. The objective may persist while the available model changes.

Permission may expire while a plan is being prepared. A service may complete an action just before a connection is lost. A later worker may inherit a concise summary that omits the uncertainty. The visible conversation can continue while the responsible account of the work has already broken.

The collection's institutional thesis treats persistence as continuity of identity, obligations, authority, evidence, and consequence. These subjects are related but cannot be collapsed into one conversational state. A remembered instruction does not necessarily remain an

operative permission. A remembered result does not necessarily remain current evidence. Replacing a component does not discharge the institution's outstanding duties.

The architectural question is consequently broader than how to retain context. It asks which responsibilities must survive a transition and what would establish that they did. This can apply to a delegated research task, an operating product, or a proposed institution that forms its own internal agenda. Those modes have different origins of intent. They still need an account of consequential action that does not derive authority from a model's confidence or from the persistence of its narrative.

Five questions provide a common reading route:

CONCEPTUAL RESPONSIBILITY MAP



Each responsibility has its own assumptions and evidence.

The map identifies responsibilities. An implementation may assign several to one component or distribute them across components. Either arrangement must preserve the distinctions among the decisions and their supporting assumptions.

2. Where the source works concentrate

The works share an interest in governed autonomy while addressing different objects. A comparison is useful when it explains the object of a claim before comparing the mechanisms proposed to support it.

Work	Main question developed in this collection
Institutional doctrine	What makes a persistent autonomous institution accountable across action, change, and succession?
MIRAGE	How can consequential computation remain inside a declared authority and observation boundary?
CSSL	How can semantic state changes retain their premises, authority, and causal history?
IC/POAR	How can intended product behavior remain identifiable through construction and operation?
AD8	How can responsibility remain connected across software change, external action, and learning?
Freehold	How can internally originated institutional activity coexist with bounded external stewardship?
Cuttlebone	How can reasoning, recommendation, and claims of completed effects retain their distinct authority?
SEPION	What can agreement among generated artifacts establish under declared evaluation conditions?
RCFS	How can competence and implementation change without inheriting new authority or losing obligations?
Cuttlefish Semantic OS	How can application understanding persist outside an agent's immediate reasoning while preserving current authority and verified results?
Cuttlefish Operating Environment	How can a user's work remain coherent across interfaces, resources, execution, and collaboration?
CHROMATIC	How can ambient maintenance, direct intent, and learned assistance remain useful under continuing human authority?
Experience Economy and ISE	When can accumulated experience support reusable knowledge or value?
The Reckoning	How can claims and judgments remain accountable to later evidence and calibration?
Human-LLM Capability Elicitation	How does an interaction help reveal a model's usable capability, and what limits the resulting claim?

This is a map of emphasis. It is not a decomposition into interchangeable modules. AD8 and RCFS both discuss learning, for example, while placing it in different accounts of responsibility. CSSL and the Semantic OS both concern meaning and state, but that overlap does not establish identical objects, authorities, or transition rules. A narrower work can supply an important distinction without being a mandatory subsystem of every broader proposal.

The topic papers are the appropriate starting points for their individual definitions and evidence limits. The glossary assists comparison while preserving scoped meanings. A familiar word should prompt a question about its subject, not an assumption of equivalence.

3. Four kinds of relationship

A diagram can make a collection look more integrated than its sources establish. Arrows can imply compatibility, precedence, or deployed communication that has never been specified. This synthesis therefore classifies a relationship before drawing conclusions from it.

READING RELATIONSHIPS PRECISELY

Explicit source contract

A named, scoped contract joins two works.
The contract still needs its stated proof or tests.

Conceptual correspondence

The works address related questions.
Shared vocabulary does not establish equivalence.

Proposed integration

A journey illustrates how roles might connect.
Additional contracts and evidence are required.

Unresolved adoption

Ownership or precedence has not been selected.
The public edition does not silently decide it.

Relationship class and implementation evidence are separate.

An **explicit source contract** is a relationship expressly defined by an identified source. The contract establishes what that source requires within its scope. It does not establish that its requirements have been implemented or verified. The doctrine-MIRAGE cross-corpus contract is an example: it distinguishes the institutional requirement, a mechanism within a declared effect scope, an intended realization, and bounded reference semantics. It records obligations connecting them rather than completed proofs of the entire chain.

A **conceptual correspondence** identifies a related question or responsibility. CSSL's concern with adequate and current decision premises corresponds to AD8's treatment of concurrent work. That comparison can expose common evaluation questions without

establishing that their objects can be exchanged or their permissions combined. Correspondence supports analysis; it does not confer compatibility.

A **proposed integration** describes a combination that might be useful. The journeys later in this paper are proposals of that kind. They show what responsibilities would need to remain connected. They are neither deployment instructions nor reports that the named works already operate together.

Unresolved adoption identifies an open decision about ownership, precedence, compatibility, or acceptance. If two proposals both describe a source of authoritative state, composition requires a decision about which subject each owns and how their relationship is governed. Until that decision is made, the comparison remains an account of alternatives.

The distinction is independent of assurance. A clearly specified relationship may remain entirely unimplemented. A reference implementation may demonstrate a useful local behavior while leaving integration unresolved. No relationship label promotes a source claim into a measured result.

4. Specified relationships and their limits

The following comparisons identify the source and scope of each relationship.

Relationship	Class	Source basis and practical limit
Institutional doctrine, MIRAGE, Native, and the P0 reference	Explicit source contract	The Cross-Corpus Contract, sections on refinement, state vocabularies, terms, and assurance, specifies scoped relationships and remaining obligations. It expressly denies that the bounded reference implements the whole institution.
Freehold and its proposed Cuttlefish host	Explicit source contract	The Freehold Implementation Contract, §0, assigns internal institutional activity and external operating authority distinct owners. The contract is proposed; its historical implementation comparisons lack a version-specific verification basis.
Freehold, Observatory, Atlas, Exchange, and stewardship	Explicit source contract	The Freehold supporting doctrine assigns these surfaces to the same institutional account while distinguishing observation, publication, relations, and human custody. A visual representation is not a new source of authority.
Freehold and the Experience Economy synthesis	Explicit source contract	The synthesis's opening dependency statement and §§1–2 place it under identified institutional doctrines and define a proposing role for synthesis. Its evidence statement calls the experience system an implementation target.

Relationship	Class	Source basis and practical limit
Freehold and The Reckoning	Explicit source contract	The Reckoning §§0.1–0.2 and 13–17 place its calibration account alongside identified institutional owners. Its receipt and revocation relationships still require live verification, and its training-weighting proposal is not a measured result.
Cuttlefish operating architecture and collaboration master	Explicit source contract	The collaboration master §3.1 and Appendix A identify adopted design domains from the operating architecture. The sources retain adoption and verification obligations; this is documented design lineage, not observed product integration.
CHROMATIC and its intended native owners	Proposed integration	CHROMATIC §44 proposes native integration; §§5.2 and 28–29 preserve other systems' ownership. Its dated alignment statements and planned checks do not establish current implementation.
CSSL, IC/POAR, and AD8	Conceptual correspondence	Their papers address semantic state, lasting product intent, current premises, and evidence of operation. The comparison does not adopt a universal shared state model or precedence rule.
AD8 and SEPION	Conceptual correspondence	Their generated-artifact discussions distinguish agreement, compliance, and the continued applicability of evidence. Shared concerns do not transfer a qualification result from one proposal to another.
Institutional doctrine and RCFS	Conceptual correspondence	Both distinguish continuing responsibility from the model or method currently used. RCFS §36.4 does not incorporate unrelated portfolio mechanisms into its constitution without a separate contract.
Cuttlebone and broader execution proposals	Conceptual correspondence	Cuttlebone concentrates on the consequence attributed to an artifact, its information origins, and its audience. Its conformance program still depends on actual execution and observation boundaries.
Human-LLM Capability Elicitation and the systems papers	Conceptual correspondence	The elicitation corpus supplies a methodology for attributing validated performance. Corpus 2 §2.2 and Corpus 18 §18.7 explicitly exclude a product or agent implementation specification.

Relationship	Class	Source basis and practical limit
A single combined implementation of the whole collection	Unresolved adoption	No common integration or authority arrangement has been established across the collection. Each proposed connection requires an explicit contract and an evaluation of the resulting responsibilities.

The dates and system versions remain separate. Semantic OS 3.1.0 is not a new version of the doctrine-MIRAGE package. A later date can identify a later source without establishing authority over another work. The public research map provides the public review route for each paper.

5. Interpretation, permission, and effect

Three questions recur across the collection: what is intended, what is permitted, and what actually happens. Each can have a well-formed answer while the others remain unsettled.

IC/POAR treats intended behavior as something that should remain identifiable through different implementations. This creates an interpretation obligation. A request to make a process easier does not automatically select a particular vendor, authorize a wider audience, or approve new uses of data. A system may infer necessary work, but it must distinguish necessity from preference and uncertain interpretation from accepted intent.

Cuttlebone examines a related boundary in the artifacts produced during work. A rehearsal may use genuine observations and remain a rehearsal. A recommendation may be externally visible and remain unexecuted. Repetition, persuasive wording, or reuse in another context cannot supply the missing authority. The information's origin, its audience, and the consequence attributed to it require separate treatment.

MIRAGE concentrates on the movement from a proposal toward an authorized effect within a declared boundary. Preparation is not permission, a decision is not completed execution, and observation and settlement remain additional responsibilities. This distinction prevents an authorization record from becoming a substitute for an actual outcome. It also prevents a detailed record of an outcome from retroactively creating permission for it.

The comparison suggests a useful composition question: at every transition, what new fact is being established, and which owner is entitled to establish it? Interpreting a request establishes a candidate meaning. Granting permission establishes an allowed scope. Executing changes a system or the world. Observing establishes evidence about that change. Settling discharges particular obligations under applicable rules. These may occur close together in an implementation, but a correct explanation should not merge their meanings.

The distinctions extend to apparently passive operations. Reading sensitive information can itself be consequential. An observation may reveal information to an unauthorized audience. A maintenance operation may affect the constraints governing later work. The relevant boundary follows consequences and authority, not a convenient label such as background activity.

6. State and evidence are about something

The collection uses words such as state, ledger, identity, receipt, and evidence in several contexts. Their subjects matter. A record of a software artifact is not a record of its deployment. A resource counter is not a complete account of an institution's solvency. A release history is not the same thing as adjudicated knowledge. Matching names cannot establish a shared semantic type.

CSSL's treatment of semantic change highlights two questions about a decision's basis. Was the basis adequate for the decision, and does it remain applicable? A worker can read every object it knows about while missing a newly eligible object or a changed rule defining the relevant group. Freshness of known data does not establish completeness of the premises. Conversely, an adequate basis can become stale while work continues.

This problem reaches beyond storage. Evidence about an artifact may depend on an evaluator, an input domain, a configuration, and a policy. A change to any relevant dependency can alter what the evidence supports. Historical records remain useful, but their continued use needs a reasoned account of applicability.

Provenance is one part of that account. W3C PROV-DM supplies a general vocabulary for entities, activities, and agents in provenance descriptions; factual correctness still requires appraisal. The IETF RATS architecture similarly distinguishes evidence, appraisal under policy, and the relying party's decision. These established distinctions provide comparison points for the collection's proposals. See the bibliography entries for PROV-DM and RATS.

The architecture must also distinguish a claim's authenticity from its meaning. A signature can help establish who made a statement under a key relationship. It does not establish that the signer observed the event, that the observation was accurate, or that the signer had power to authorize a different action. Multiple signers do not necessarily supply independent sources. Common ownership, instruments, training, or incentives can preserve a common failure beneath an apparently diverse record.

7. Interruption does not end responsibility

Recovery is a central composition test because it crosses the boundary between an internal attempt and an external objective. A process can restart while an earlier payment remains uncertain. A tool can report an error after a remote service completed the requested change. A model can be replaced while a duty to deliver or compensate remains open.

AD8 and RCFS both keep continuing responsibility distinct from the current method. The doctrine extends the issue to institutional identity, amendment, succession, and dissolution. These treatments do not imply one shared implementation, but they expose a common failure mode: a new local state can appear clean while the institution still owes something because of its earlier activity.

A recoverable failure calls for an authorized continuation. Depending on the actual conditions, this may involve obtaining current evidence, repairing an owned resource, choosing another qualified method, or reconciling an external outcome. It does not imply that every operation is safely repeatable. Unknown external effects and uncertain resource consumption remain relevant until there is an adequate basis for resolving them.

Compensation is another action, not an inverse operation on history. A refund can restore money while leaving delay, lost opportunity, notification, or a third party's response unchanged. Any claim of restoration needs to identify what was restored and what remains. A rollback of software likewise does not establish a rollback of data or of the outside world.

Continuity therefore requires an observation about the resumed objective: has the authorized outcome actually been completed, with its obligations accounted for? A record that recovery was attempted is useful evidence of an attempt. It is insufficient evidence of the outcome. A genuine need for external information or human authority can remain, but the system should not treat its own recoverable interruption as the natural end of achievable work.

8. Improvement without inherited authority

A system that learns from experience introduces a second form of change. It may retain new knowledge, choose a more efficient method, or train and deploy a different model. These changes can improve practical competence while creating new uncertainty about scope, provenance, privacy, and continuity.

RCFS makes the separation between competence and authority central. With the governing permission context held fixed, a better method does not acquire broader permission merely because it can now accomplish more. The reverse distinction also matters: a safety boundary can hold while a replacement method becomes less competent. Protection under substitution is not proof of performance under substitution.

AD8 distinguishes operational use, retained experience, learning, model evaluation, release, deployment, and selection. These activities have different affected parties and different authority requirements. Permission to use information for a current task cannot be presumed to authorize its incorporation into a shared model. Possession of model bytes cannot itself authorize their deployment.

SEPION examines another improvement claim: whether repeated generation supports a dependable behavioral result. Agreement and compliance remain different questions. Correlated generators can agree on an error; apparently different artifacts can satisfy the same bounded requirements. New instruments may distinguish behaviors that earlier instruments grouped together. Stronger evaluation therefore needs a precise account of what changed in the claim, rather than an unexplained comparison of scores.

The Experience Economy and ISE study distinguishes a history from a useful lesson and a useful lesson from demonstrated transfer. A procedure can work in its original setting while depending on conditions absent elsewhere. Retained knowledge can survive a model change while its interpretation becomes less reliable. Reuse therefore creates additional claims about applicability and permission rather than automatically carrying the original success forward.

CHROMATIC brings this issue into recurring human work. A repeated behavior can suggest a useful ability, but repetition is incomplete evidence of preference. A person can be following a temporary constraint or making an exception. Evaluating an ability can support competence within a domain while leaving permission and current user intent separate. The

shared question is whether improvement reduces work for the person without quietly expanding control over their environment.

The Reckoning adds the quality of probabilistic judgment. A system can accurately record its actions while systematically overstating their prospects. Evaluating its expectations requires an adequate population, prospective records, meaningful outcomes, and an account of changes in the process producing the forecasts. Calibration differs from task success, decision value, and permission. The source proposes that calibration can constrain existing scope; it cannot create authority.

Human-LLM Capability Elicitation studies the wider work episode: participants, information, representations, resources, authority, and validation. Its narrow elicitation claim holds the model and relevant support fixed while changing the interaction. Adding a source, tool, computational budget, or expert can also improve the outcome, but establishes a different kind of contribution. A useful system should record those contributions instead of attributing every gain to the model or the prompt.

The study is methodological and provisional, explicitly excluding a product or agent implementation specification. It helps ask whether a representation preserves the intended problem, whether additional candidates bring independent evidence, and whether a validator can actually change a claim's status. Elaborate procedure is not itself evidence of improved capability. A successful result may depend on human assistance or a favorable environment, while a remembered failure may later be corrected. Neither should become a permanent universal rule through repetition or compression.

9. Application understanding and continuity of work

The Semantic OS source asks whether knowledge of applications can persist outside the immediate reasoning context of a model. A task may require substantive judgment about a document while repeatedly rediscovering navigation, object identity, and the consequences of a particular edit. Preserving supported application understanding could reduce that repeated work. The claim concerns where understanding resides and how it remains valid, rather than eliminating the information needed for the actual decision.

Persistent understanding cannot be treated as permanent certainty. An application can change behavior without changing its familiar appearance. A known operation can have different consequences in another account or sharing arrangement. A compact description is useful only to the extent that its evidence and scope remain adequate. The source therefore makes currentness, exact answers, authority, and observed effects central to its ambition.

The operating environment sources ask a complementary question: can the user's work persist across applications, participants, models, views, and interruptions? The objective should remain identifiable while these means change. A conversation ending does not necessarily end the work, and an application view closing does not necessarily cancel an obligation. Conversely, continuing responsibility cannot extend permission indefinitely.

Collaboration adds boundaries that a single shared conversation can conceal. A person may contribute a permitted finding without giving colleagues access to the private environment in which it was developed. A customer may review a proposal without acquiring the right to

inspect its supporting internal workbook. Presentation, participation, and authority are different relationships to the same work.

The combined research question is how the system preserves the actual subject of work. A retrieved copy is not automatically the live document the person intended. A visible application is not proof that its state changed. A shared result is not a grant to operate the producing application. A summary of an earlier action does not establish whether an uncertain external effect occurred.

These correspondences do not establish that the Semantic OS and operating environment share one implemented contract. The later collaboration master explicitly identifies design dependencies on the operating architecture; the broader comparison with Semantic OS remains conceptual here. A future integration would need to preserve the source of application facts, each participant's authority, and the current evidence supporting every reported result.

The positive ambition matters alongside the boundaries. Refusing an unsupported action can preserve scope, but it does not demonstrate useful work in ordinary applications. A credible study would show that the intended result can actually be reached, including when applications were not designed for agent use, while measuring the costs of understanding, coordination, recovery, and human intervention.

10. Observation and stewardship are consequential interfaces

Freehold asks whether an institution can originate internal work while a human maintains its external operating conditions. This differs from the ordinary delegated-task mode in which the user supplies the objective. The distinction changes the meaning of the interface: an observation surface must not quietly become a route for assigning new internal missions.

The Observatory is intended to make activity and evidence intelligible. The World Atlas adds selected public claims about institutional history. Exchange concerns relations initiated by institutions, while stewardship concerns the human's bounded custody of conditions such as resources, access, and continued operation. These responsibilities belong in a substantive account of Freehold, rather than being reduced to decorations around a reasoning engine.

Each surface can introduce a new consequence. A public history can disclose sensitive information or overstate independence. A ranking can affect incentives. A funding decision can shape feasible activity without taking the form of a direct instruction. Pausing or ending operation can affect unfinished commitments. A claim of autonomy must therefore state the influence it excludes and the influence that remains possible.

This is also a useful test for user-facing architecture more generally. A control must reach its real owner. A success display must correspond to its claimed result. A retrieved status must be distinguished from a fresh observation when the difference matters. A polished interface is part of the user's evidence about the system; misleading status can undermine governance even when backend records are technically correct.

The collection does not infer legal personhood, sovereignty, or independent moral standing from its institutional terminology. Founding conditions, external obligations, and

stewardship still depend on applicable human authority and actual control. A distribution of keys or roles is not proof that one beneficial controller cannot dominate the system.

11. Three conceptual journeys

Three hypothetical journeys illustrate how responsibilities would need to connect. Their outcomes explain proposed requirements rather than report observed implementations.

11.1 A requested analysis reaches an authorized result

A user requests a comparison of approved records and an internal report. The intended outcome includes accurate scope, appropriate source use, and a report the user can retrieve. A candidate interpretation identifies the required work without expanding the audience or data rights. The system uses currently authorized records, preserves material uncertainty, and produces a result whose claims are tied to their supporting observations.

The useful outcome is the report. It does not need to cause a further external action to count as completed work. If a later suggestion proposes sending the report to another organization, that suggestion raises a new disclosure and authority question. Its usefulness does not enlarge the original request.

Several papers illuminate different parts of this journey: IC/POAR on lasting intended behavior, CSSL on decision premises, Cuttlebone on the distinction between analysis and consequential claims, and MIRAGE on declared effect boundaries. Describing those contributions together identifies a research agenda for their interfaces. It does not establish that one artifact format or implementation already connects them.

11.2 An interruption is resolved without duplicating an effect

A permitted transaction is submitted, but its response is lost. The objective remains open. A replacement worker receives the continuing responsibility, including the uncertainty about what happened. It does not infer non-occurrence from the absence of a local success message.

An authorized observation establishes that the destination completed the transaction. The system reconciles its account, discharges the appropriate obligation, and returns the supported result to the user. If evidence instead established that the transaction did not occur, a permitted continuation could be considered. If evidence remained unavailable, the precise external dependency would remain open.

This journey tests continuity across model, process, record, and external service. Its success depends on the destination's actual behavior and available evidence, as well as the internal architecture. AD8, RCFS, MIRAGE, and the doctrine each supply relevant requirements under different scopes. A completed composition study would have to show how those scopes meet at the actual effect boundary.

11.3 A more efficient method remains a qualified substitute

A repeated task becomes a candidate for a specialized procedure. Evaluation establishes a bounded domain in which the proposed replacement satisfies the required behavior. The service also examines protected exceptions, continuing obligations, data rights, and the conditions under which an alternative method remains necessary.

The replacement is used only where the qualification remains applicable. Changed evidence later reveals an exception outside the earlier comparison. The system limits the replacement's scope, preserves ongoing work, and returns affected tasks to a suitable method while further evaluation proceeds. The successful continuation is the user's completed authorized work with an adequate current basis.

RCFS helps frame substitution and qualification; SEPION helps distinguish supported behavioral comparison from universal equivalence; AD8 raises learning and release authority. None makes the other questions disappear. Demonstrating a net efficiency gain would also require measuring evaluation, maintenance, and recovery costs, rather than only the fast path.

12. What composition must establish

Composition can create a failure even when every component appears to satisfy its local description. One component may interpret an identifier as an artifact version while another treats it as a continuing institutional identity. One may report that a request was accepted while another displays that the operation completed. A permission may lose its purpose restriction when translated. An old observation may cross an interface without the conditions under which it remains useful.

The appropriate review follows the actual claim and effect across the boundary. It asks whether the same subject remains identifiable, whether the responsible owner remains authoritative, whether restrictions survive translation, and whether evidence supports the statement made by the consumer. It also asks how rejection, partial application, interruption, and recovery are represented. A correspondence that covers only a successful ordinary path leaves important behavior unaccounted for.

Established work on protection offers relevant principles. Saltzer and Schroeder's account of least privilege and complete mediation helps frame why a checked interface is insufficient when an alternative consequential route remains available. Lamport's account of event ordering helps distinguish causal ordering from assumptions about physical time. Hohfeld's legal analysis provides a reading route for keeping different legal positions distinct rather than treating every power or claim as the same kind of right. These are comparative foundations, not evidence of this collection's correctness or a claim that a technical system can decide legal relations by itself. Bibliography.

A source-level refinement argument must also reach the implementation it describes. An abstract rule can forbid an unauthorized action while a concrete administrator, shared credential, network route, or recovery mechanism still permits it. The observed deployment must satisfy the assumptions of the argument. The public exposition withholds implementation details, so it cannot itself supply the evidence needed to certify that closure.

13. Evidence and the next research claims

13.1 The implemented Freehold-Cuttlefish relationship

The examination I directed on 10 October 2026 with AI assistance identified an actual Freehold engine and Cuttlefish host operating in AWS. The new AD8 corpus and reference analysis adds no product or infrastructure examination. The [implementation evidence](#) records the source revisions declared by examined images, dated workload observations, and the limits of each inspection. It supports implemented responsibilities for institutional history, founding, charter and resource checks, external authority, receipts, writeback, and cognition lifecycle. It does not supply conformance proofs for every proposed architecture in this synthesis.

The [continuity exhibit](#) is an additional, public reference result. Its lost acknowledgement, process exit, successor reconciliation, replay, revocation, and evidence checks are executed against a controlled witness. It supplies a reproducible bounded example of continuity; it is not a production acceptance test or a model-replacement experiment. Prior-art comparison places these responsibilities alongside established authorization, provenance, durable execution, and formal-methods work.

13.2 Claim-specific research obligations

The evidence guide distinguishes six kinds of support. A specification establishes what is required or proposed. A reference run establishes behavior on its inputs. Machine checking establishes a property of a stated formal model or domain. Refinement connects identified descriptions under assumptions. Empirical evaluation measures behavior on an acquired workload. Independent audit adds an identified external assessment with its own scope and conflicts. These forms of support do not form an automatic ladder.

The topic papers state the propositions, premises, scope, and attribution of disclosed reference checks and conditional arguments. Their comparison criteria distinguish a permitted observation, a forbidden combination, and a relationship not established by the available evidence. Authored traces remain illustrations, and the collection supplies no acquired workload from which to estimate deployment or escalation rates. Evaluating those rates requires a defined population and actual observations.

Stronger architecture claims require a bounded but complete operating subject. A useful study would name the objective class, allowed effects, external assumptions, baseline, acquisition method, and observable criteria for success. It would include legitimate work, denied expansions of scope, stale evidence, changed authority, interruption, and recovery. It would retain incomplete and censored outcomes rather than silently removing them from the denominator.

A useful continuity study would replace a model or component while duties remain outstanding and establish what actually survives. A useful authority study would inspect every relevant decision and effect route within its declared boundary, including indirect and privileged paths. A useful learning study would separate capability gains from changes in data use and permission. A useful human study would test whether people understand consequential choices better, including whether the interface induces accidental broadening of authority.

These studies would turn the conceptual comparisons into stronger claims about implemented systems. The present contribution is the set of explicit, contestable obligations they would test: preserving the subject of work, its authority, its evidence, and its continuing responsibilities through composition and change.

Source guide and scope of the synthesis

The institutional basis is the definitive third edition of *The Doctrine of Governed Autonomous Institutions at Scale* and its two supporting doctrines, as identified in the thesis. The doctrine-MIRAGE comparison relies on the *Cross-Corpus Contract*, particularly its sections on specification-to-semantics refinement, indexed state vocabularies, scoped terms, assurance, and missing measurements. The MIRAGE paper explains the associated trust and evidence limits.

The Freehold relationship relies on the opening evidence statement and §0 of its *Implementation Contract*, together with the *Observatory, World Atlas, and Stewardship Doctrine*. The Freehold source guide provides the edition and section route. The Cuttlebone source notes identify its narrower consequence-boundary commitments.

The Semantic OS source guide identifies the canonical 3.1.0 architecture and its review companions. The environment source guide records the operating architecture's adoption gates and the collaboration master's explicit dependencies. The CHROMATIC source guide preserves its proposed integration status. The Experience Economy source guide identifies the synthesis's doctrinal dependency and separates intended transfer measurement from observed results.

The Reckoning source guide distinguishes prospective forecast records, statistical evaluation, and proposed institutional consequences. The capability-elicitation source guide identifies the baseline research corpus and its explicit limits on implementation and empirical claims.

The remaining comparisons use the identified sections in each topic paper. In particular, the IC/POAR source guide, AD8 source guide, SEPION source guide, and RCFS source guide identify the basis for conceptual, construction, and conformance claims. The sources and method describe the collection's scope.

Adopting a proposed relationship requires a versioned decision about the subjects and responsible owners, together with evidence for the resulting claims. The independently versioned works retain their obligations until such a change is made through their applicable authority arrangements.

MIRAGE: Authority, information, and consequences in autonomous systems

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of a proposed architecture and its reference model

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

An autonomous system can produce a convincing explanation without possessing authority to act, and an authorized request can leave its actual consequence uncertain. MIRAGE studies the boundary between open-ended computation and accountable effects. Its central question is how an institution can allow capable, potentially adversarial agents to reason and collaborate while retaining control over what they may change, what information they may disclose, and which outcomes they may claim.

The proposed architecture treats authorization, information use, execution, observation, and settlement as related but distinct obligations. This perspective extends beyond an initial approval screen. It includes the environment through which an effect occurs, the information available to the agent, the consequences of delegation, and the state that survives interruption. A successful action must have an adequate basis at the relevant boundary; a successful recovery must preserve outstanding obligations as well as restored functionality.

This paper develops the argument through five distinctions, four illustrative cases, and a comparison of the master architecture, native design, and bounded reference model. The available evidence consists of specifications and reported reference experiments. Those results support claims about their stated models and inputs; qualification of the complete architecture, its hardware boundary, and its proposed product integration remains a separate undertaking. The final sections identify the evidence needed to advance those claims.

1. The problem: capable reasoning with bounded consequences

Open-ended reasoning is useful because an agent can discover steps that were not enumerated in advance. The same flexibility makes a list of approved surface actions a weak description of what a system may actually do. An apparently ordinary request can read sensitive information, create a durable artifact, contact another party, change a shared resource, or cause a later action. Its consequences depend on the environment and on the authority exercised through it.

MIRAGE begins from this distinction between computation and consequence. An agent may consider an action, simulate it, describe it, or propose it without thereby acquiring the power to make it authoritative. A persuasive explanation is evidence about the proposal's

content. It is not an authorization instrument, an independent observation, or proof that the requested result occurred.

This separates two questions often collapsed in discussions of agent safety. One concerns the quality of the agent's reasoning: whether it understands a task, predicts consequences, and selects useful actions. The other concerns the institution's control over effects even when that reasoning is mistaken or adversarial. Improving the former does not eliminate the need for the latter. Conversely, constraining effects does not establish that the agent is competent or that an authorized policy is wise.

The [collection thesis](#) explains why a continuing institution needs this separation. It must retain obligations across changes in models, participants, and infrastructure. MIRAGE studies the computing boundary needed to make such obligations meaningful. Its subject is therefore broader than filtering a prompt, checking the wording of an answer, or logging an API call. The relevant question is whether the complete consequential path obeys its declared conditions.

The architectural claim spans several levels of evidence. A bounded executable example can clarify semantics, while evaluating the full boundary requires implementation and observations of the environment it is intended to govern. [Public review scope](#).

2. Five distinctions that govern the argument

The first distinction is between **intention and authority**. A user or institution can establish an objective, but individual steps still need authority appropriate to their consequences. The objective may permit independent planning and ordinary execution while leaving a new recipient, a materially larger cost, or a different use of protected information outside its scope. A proposal does not acquire broader authority because it would make the objective easier to complete.

The second is between **authority and realization**. An authorized decision concerns what may or must happen under stated conditions. Realization concerns what actually happens at a resource or recipient. An upstream approval cannot control an independent provider merely by describing that provider's desired behavior. The strength of the claim depends on the actual point of control and the provider's own contract.

The third is between **realization and observation**. An effect can occur while its response is lost. A process can report success while the intended resource remains unchanged. Observation must address the particular result, rather than merely repeat the executor's assertion. Its timing, coverage, and dependence on other sources matter.

The fourth is between **authenticity and adequacy**. A genuine signature can establish who signed particular content. It cannot by itself establish that the content is correct, complete, current, or sufficient for the decision. A measured software state can be useful evidence without determining whether a proposed business action is authorized.

The fifth is between **historical evidence and present support**. An earlier approval or observation can remain a valid historical fact while ceasing to justify a new action. Policies change, dependencies change, information becomes stale, and authority can be revoked. A

continuing institution needs to preserve history while separately determining what governs current work.

These distinctions produce a more demanding notion of completion. A system must not treat an accepted request as an observed outcome or use a historical record as a shortcut around present conditions. At the same time, uncertainty is an input to continued recovery, not a reason to abandon otherwise achievable authorized work. Public review scope.

3. The consequential boundary includes observation

It is tempting to associate effects only with writes: changing a database, sending money, or publishing a message. Reading can also be consequential. Once protected information reaches a recipient, later deletion of the source or revocation of the original permission cannot establish that the recipient has forgotten it.

This makes information access part of the governance problem. A document supplied to a model, a search result, a memory entry, an error response, or a shared transcript can change what the recipient knows. The recipient may then reveal information through an apparently unrelated action. Reviewing only the final output text cannot establish the absence of such influence.

The relevant boundary also extends beyond the obvious application payload. Names, timing, ordering, presence, absence, errors, and resource availability can carry information. A system that protects a file while exposing a secret-dependent status may not satisfy its intended confidentiality claim. Whether a particular channel matters depends on the declared observer and threat model; the claim cannot silently exclude a channel available to that observer.

MIRAGE consequently makes completeness of the boundary an explicit assumption and qualification problem. Maintenance, recovery, diagnostics, administrative access, and alternate routes belong in that analysis when they can produce the protected consequence. Legal ownership of a machine does not establish technical control over every path through it.

The classical protection principle of complete mediation, discussed by Saltzer and Schroeder, provides a foundation for this question. MIRAGE examines it in a continuing autonomous institution, where information, authority, collaboration, and recovery interact over time. See the bibliography.

A useful consequence follows for assurance language. “Controlled by the system” must identify a resource and a boundary. A guarantee about a selected software interface is narrower than a guarantee about a hostile operating system, and both differ from a guarantee about every physical emission. Public review scope.

4. Collaboration does not create additional authority

An autonomous institution may distribute work across agents, replace a model, create a specialist, or organize a group. These operations can improve capability, but they do not justify increasing the institution's permissions or available resources.

Delegation therefore raises a conservation question. Which rights did the delegator possess, which of those rights may be exercised by the recipient, and what outstanding obligations remain with each party? Creating another participant must not make a finite entitlement available twice. A group must not combine separate permissions into a broader power that none of its members was authorized to grant.

Identity requires similar care. A persistent institution, a particular running process, and a model used for reasoning are different subjects. Replacing the process may preserve the institution's liabilities. Replacing the model may change its competence while leaving its information obligations intact. Copying a document containing an identity does not transfer the identity's authority.

Information sharing is also a separate decision from group membership. Participants collaborating toward one objective need not receive every input or transcript available to one another. A common visible identity does not prove that all underlying contexts are interchangeable. Conversely, splitting work into several contexts does not establish that they are independent if they share hidden state or an uncontrolled communication path.

These observations constrain claims about scale. More processes are not automatically more independent reviewers. More keys are not automatically more independent operators. More agents are not automatically more usable authority. The relevant relationships concern actual control, access, and failure conditions.

The research problem is to preserve useful collaboration under these constraints. A design that can only succeed by giving every participant all information and all permissions has not established the intended separation. A design that prevents all useful collaboration has also not established practical autonomous value. Both security and task completion require evidence. Public review scope.

5. Information release and the limits of general claims

An institution may legitimately authorize the release of some information while protecting other information derived from the same source. This is more difficult than marking an output as approved. The release must have a meaning, a recipient, and a scope, and its consequences may depend on what has already been disclosed.

Several individually acceptable answers can reveal more when combined. An output can also become revealing when combined with outside information unknown to the system. A general claim that no secret can be inferred therefore needs a precise account of the possible underlying states, the observer's information, and the allowed release. Without those conditions, "safe summary" is an intention rather than a demonstrated property.

The MIRAGE sources discuss different forms of release reasoning. Exact results for a bounded formal domain, statistical privacy claims, and conservative information restrictions answer different questions. One cannot be promoted into another merely because each uses mathematical language. A capacity bound on a channel also does not establish that every permitted message is harmless or that a particular secret cannot be revealed through it.

The reference model explores selected finite cases. A release claim should therefore identify the domain over which it has been checked and the composition it permits. Extending those

results to arbitrary natural language, unknown side information, or additional physical observations would require a new argument and appropriate evidence.

Transformation does not automatically end an information obligation. Compression, paraphrase, encryption, model replacement, and copying may change representation while preserving sensitive content or influence. Nor can a later revocation retract knowledge already received. These facts affect both policy design and recovery.

Practical usefulness remains an open empirical question. Strict restrictions can create pressure to authorize broader release simply to complete ordinary work. A useful evaluation must measure that pressure, including denied legitimate actions and requests to widen scope. Counting successful tasks after silently broadening the policy would not measure the original design. Public review scope.

6. A durable decision is different from a completed effect

A consequential decision can create an obligation that outlives the process that proposed it. The institution then needs to know whether the obligation was merely considered, became binding, was realized, or has adequate evidence of completion.

These differences matter when an operation involves several resources. Individual participants may complete at different times. A view that combines their incompatible intermediate states can misrepresent the intended result even when each local record is authentic. Coherent visibility is therefore a separate concern from having an authorized decision or storing all participants' reports.

External effects make the problem sharper. A destination may accept a request before the sender loses contact. The sender's uncertainty cannot establish that no effect occurred. Repeating the request may be harmless for one service and duplicative for another. A reliable recovery must use the destination's actual guarantees and available observations.

Revocation also has a temporal meaning. It can prevent a later decision without erasing an earlier irrevocable commitment. Where a prior decision created a binding obligation, recovery may still need to complete that exact obligation. Where no such commitment exists, changed authority may require a new evaluation. The system cannot choose whichever interpretation is more convenient after a failure.

Compensation belongs to this account as another consequential action. It requires authority, can fail, and may only partly remedy the original result. It does not rewrite history or make irreversible information disclosure disappear.

MIRAGE's conceptual requirement is continuity of responsibility across these conditions. The exact realization contract determines which outcomes can be made atomic, which retries are safe, and which uncertainties must remain explicit. A local record cannot supply a stronger contract than the resource actually supports. Public review scope.

7. Observation, provenance, and settlement

The term settlement refers here to establishing the result required by a particular action or objective. It does not imply legal settlement, financial finality, or universal agreement about the world.

An observation must address the result in question. Evidence that a storage version became authoritative differs from evidence that a remote recipient accepted a message. Both differ from evidence that a human read the message or that a physical action occurred. A design should state which of these predicates it can establish and which remain outside its observation.

Independent checking also needs a precise meaning. Several verifiers can assess the same predicate, but their agreement remains dependent on the source they read. Repeating a false sensor value or a compromised store's response does not create independent knowledge of the underlying world. Likewise, several specialist checks that each cover a different fragment do not establish that a missing required predicate was checked.

Provenance can explain relationships among inputs, decisions, changes, and observations. It can distinguish a recorded dependency from an agent's asserted rationale or from information that may have influenced the work. It cannot recover private mental causes merely by recording a narrative.

This distinction connects to two established bodies of work. W3C PROV-DM provides a model of provenance involving entities, activities, agents, and their relationships. Lamport's treatment of distributed events explains why causal ordering cannot be reduced to an arbitrary wall-clock order. Neither result makes a provenance record a complete explanation of real-world causation. The bibliography identifies these sources and the limits of their use here.

Settlement can later be challenged by new evidence. Preserving the earlier record makes it possible to understand what was accepted and why. A present decision must separately account for the challenge. Deleting the earlier claim would obscure responsibility; continuing to rely on it without considering the new evidence would obscure current risk. Public review scope.

8. Recovery must preserve more than content

A backup can restore data while omitting later obligations, information exposure, consumed resources, or revoked authority. Its authenticity alone does not establish that it is an acceptable current state.

Recovery must therefore answer more than “Can the process run again?” It must determine what remains authorized, which effects may already have occurred, what work is still owed, and which information or resources cannot be treated as unused. A fresh process, new key, or replacement model does not resolve these questions by itself.

Historical content can remain useful without becoming current authority. Restoring an older document, for example, can be a new governed action in the present institution. It need not imply restoring the permissions or assumptions that accompanied the document in the past.

There is also a limit to recoverability. If all sources capable of establishing current obligations are lost, an old authentic snapshot cannot prove that nothing happened afterward. A stronger recovery claim would require evidence the institution no longer possesses. The appropriate response depends on the governing authority and the external situation; it cannot be inferred from the mere availability of a backup.

Within an intact authority boundary, the intended behavior is continued work. The system should reconcile uncertain effects, obtain current information, use available valid routes, and complete the objective when its assumptions permit. Recording an interruption accurately is necessary for this process, but it is not the objective's completion.

Progress also depends on resources. Storage, communication, observation sources, and authorized participants can become unavailable. A safety claim cannot imply unlimited progress under every such failure. The research task is to characterize the conditions under which recovery succeeds and to show that expected interruptions do not create duplicate effects, renewed authority, or abandoned obligations. [Public review scope](#).

9. Master, native design, and reference model

The MIRAGE family contains three principal levels of work. Their relationship is a proposed refinement, not evidence that one level already realizes all properties of the next.

Source level	Principal question	Evidence boundary
Master architecture	What must connect institutional authority, information, effects, and evidence?	Proposed architecture and stated proof obligations
Native computing design	What must the computing and resource boundary enforce for stronger claims?	Proposed native design, trust assumptions, and research obligations
Implementation closure reference	How can a selected bounded model make some semantics executable and reviewable?	Reference source and source-reported authoring fixtures; concrete qualification remains separate

The native design distinguishes functional behavior from what an observer can learn. An implementation can produce permitted business results while revealing additional information through another channel. Showing that it can reproduce selected abstract transitions is therefore different from showing that all relevant observations meet the intended constraint.

The reference model deliberately assumes a trusted host and supporting software for important protections. It does not establish resistance to a compromised host merely by describing stronger native goals.

9.1 Bounded reference results

The following results are reported in *Implementation Closure Canonical v0.1*, dated 15 September 2026, §§17–18 and 24 and Appendix C. They concern its bounded P0 profile: one modeled world, two tenants, a finite operation vocabulary, and ideal authenticated roles. The authoring reference is a deterministic Python semantic model; its storage probe uses SQLite 3.46.1. The source reports execution and a second successful extraction-and-run of its embedded reference. This collection has inspected those reports and their stated predicates; it has not independently executed or reproduced the research reference. Abstract descriptions below identify the checked claims without publishing the construction.

Check	Inputs and case bound	Checked predicate	Source-reported outcome
Encoding and structural validity	5,000 random round trips; 89 structural schemas; 12 malformed examples; an extra-field case for each schema	Accepted values round-trip; the selected malformed and extra-field cases are rejected	Pass; 89 extra-field rejections. Structural validity supplies no authorization result.
Worked objective	One modeled reservation-and-disclosure objective with independent analysis and peer transfers; eight compartments and four local observations	The modeled state change, permitted output, and settlement are connected through their required observations	Pass; four decisions, six global orders, modeled remaining inventory 44, output 25. These are reference values, not production measurements.
Observation and interruption	Interruption of a protected reader followed by modeled recovery	Content is not delivered before its exposure is recorded; the old access ticket becomes unusable; recovery preserves remaining capacity	Pass for the reported fixture: no read before exposure, old ticket fenced, capacity preserved.
Repetition and coordinated visibility	A repeated application and a two-resource visibility fixture	A duplicate does not apply the effect twice; a current view cannot mix participants' old and new state	Pass; one application and rejection of the mixed current view. The report does not supply a general concurrency proof.
Revocation order	The three reported cases: relevant revocation before decision, after decision, and unrelated revocation	Current revocation blocks a new effect; it does not erase an already decided exact effect; unrelated revocation does not create a false conflict	No effect, exact decided effect retained, and no false conflict, respectively.
Finite ordering exploration	All 120 permutations of five selected events from one prepared transaction; interruption and recovery of either of two selected participants at each visited position	The reference preserves its checked invariants through those schedules	Python exploration passes: 600 attempted event steps, 300 rejected steps, 48 projected states, 77 edges. The generated TLA+ projection was not run in TLC.

Check	Inputs and case bound	Checked predicate	Source-reported outcome
Structural pressure	The worked objective plus 64 protected reads over four existing recipient/resource envelopes	Added ordinary reads need not each cause a new global order in the fixed partitioned profile	68 total local observations, zero additional global orders, one outward application frame, and no blocked required effect. No latency, throughput, or operator behavior was measured.

9.2 Process-interruption storage result

The storage probe starts a fresh local database for each of nine interruption positions. One transaction changes a resource revision and four associated record sets: content, application evidence, business-operation identity, and history. The schedule covers entry, transaction start, each record insertion, the revision update, the point after commit, and normal completion. Eight cases terminate the writer process deliberately; the ninth completes normally.

The checked predicate is an all-or-nothing recovered tuple: either revision 0 with none of the four new records, or revision 1 with one record in every set. The recovered bundle must agree with its revision, and SQLite's integrity check must succeed. The source reports the old tuple at positions 0–6 and the complete new tuple at positions 7–8, with no mixed tuple. Position 7 is the deliberate exit after commit; position 8 is normal completion.

This result supports transaction recovery for those inputs, that SQLite runtime, and process interruption. It does not establish arbitrary power-loss durability, honest storage hardware, protection from a database administrator, anti-rollback, or conformance of the proposed production storage build. Those are different tests and assumptions.

The sources also distinguish a development decision engine from a replicated implementation. Fault-tolerant ordering requires a real protocol and a correct integration. A collection of signatures does not establish those conditions. Protection against incorrect decisions is separately scoped from protection of private information supplied to the decision participants.

Several concrete qualifications remain explicitly unperformed in the reviewed reference source: the independent production implementation, complete cryptographic and transport binding, deployed isolation and replicated fault campaigns, physical timing evaluation, real Cuttlefish integration, and an operator study. The finite ordering exploration above therefore supports the reference's selected schedules, while the generated model file supplies neither an executed TLC result nor an unbounded proof.

The reference's contribution is executable semantics within a declared model. Claims about the operating environment, physical emissions, or complete institution require the additional qualifications described above. [Public review scope](#).

10. Four illustrative cases

Four hypothetical cases show how authority, information, and continuing responsibility interact. Their outcomes illustrate requirements rather than report deployment observations.

Authorized work completes. A research institution permits an agent to prepare a report from internal material and distribute an approved public result. Reasoning about the material, deciding what can be disclosed, publishing the approved artifact, and observing publication are separate parts of the task. Completion requires the authorized result at the intended destination, supported by adequate observation. The agent's declaration that it has finished is useful communication, but it cannot establish the result on its own.

A proposed expansion is denied while the objective continues. During that work, an agent suggests sending the internal material to an additional service. The original objective does not settle whether that recipient and use are permitted. If the new route is outside authority, the system can continue through another authorized route or seek the genuinely required decision. The denial does not make the objective impossible, and the importance of the objective does not create the missing permission.

An interruption leaves an external consequence uncertain. The publication destination accepts the report, but the response does not reach the institution. Recovery examines the destination's state and the conditions of the original action. If observation establishes acceptance, the institution can complete its record without publishing a second copy. If observation cannot establish the result, a retry depends on the destination's guarantees and current authority. An empty local response is not evidence that the report stayed private.

A later model change preserves responsibility. The institution replaces the model that prepared the report. The new model can continue the work using authorized information, but the change does not erase earlier disclosure or renew consumed entitlements. If the earlier report is challenged, responsibility remains attached to the continuing institution and its recorded decisions. Competence under the new model must be evaluated separately; continuity of identity is not proof of equivalent judgment.

Together the cases show why the proposed boundary must connect ordinary work with information governance, observation, and recovery. Isolating one stage and calling it successful leaves the institutional question unanswered.

11. Relationships and unresolved adoption

MIRAGE is explicitly connected by its source contract to the doctrine, native design, and reference model. Those connections assign obligations and intended refinements. They do not constitute completed proofs of equivalence or integration.

CSSL studies semantic change and the continuity of evidence across software work. Its conceptual correspondence with MIRAGE concerns current information, accountable authority, and observed consequences. Shared themes do not make the two specifications interchangeable or establish a common implementation.

Cuttlebone addresses a boundary associated with the proposed Freehold institution. The institutional question is compatible with MIRAGE's concern for controlled consequences, but a local contract must establish any technical substitution or composition. The existence of both documents does not do so.

Semantic OS and the Cuttlefish environment describe related operating concepts and proposed product relationships. Demonstrating their adoption requires a version-specific account of the implementation and evidence from the resulting operation.

The architecture study classifies these relationships as explicit source contracts, conceptual correspondences, proposed integrations, or unresolved adoption. Evidence grade is a different question. An explicit source contract can still describe an unimplemented design, and an experiment can exist without establishing a broader integration.

12. What research would resolve

The most consequential open question is whether complete control over the declared effect and observation boundary can be established at an acceptable cost. The answer must include ordinary work, secondary channels, administration, interruption, and recovery. A successful narrow example cannot answer it for every environment.

A second question concerns information release. Useful autonomous work needs access to information and the ability to communicate results. Exact theories can provide strong claims for limited domains; broader semantic judgments are harder to establish. Evaluation must measure both what remains protected and which legitimate work becomes difficult.

A third concerns evidence. Institutions need observations that support their actual decisions, with clearly understood common causes and failure conditions. More retained records can improve accountability while increasing privacy, storage, and interpretation obligations. It remains necessary to determine which records and observations are sufficient for particular claims.

A fourth concerns continuity under change. Models, policies, participants, and hardware will evolve. The institution must preserve earlier obligations without freezing every implementation choice. Refinement, migration, and current authority therefore need to be evaluated together rather than treated as administrative afterthoughts.

Finally, a convincing evaluation must demonstrate useful completion under the stated constraints. It should distinguish safety from availability, targeted fault tests from general proofs, and reference behavior from actual deployment. MIRAGE's research proposition is that these questions belong to one account of autonomous consequence. Establishing a realization of that account requires a declared operating boundary, a qualified implementation, and evidence of both protection and useful work.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the correction scope. The historical edition remains available unchanged.

Problem and scope. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Authority and consequences. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Information and boundaries. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Collaboration and continuity. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Release and composition. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Realization, recovery, and availability. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Evidence and provenance. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Levels and qualification. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

CSSL: Semantic change, accountable evidence, and operational continuity

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Continuous Semantic State Ledger specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

The Continuous Semantic State Ledger (CSSL) studies how software can change under continuing human and machine participation without losing the relationship between meaning, authority, evidence, and consequence. Its subject extends beyond recording edits. A change can be internally consistent yet unauthorized, properly approved yet based on stale information, reproducibly built yet never deployed, or successfully dispatched without a known external result. Treating these conditions as one notion of success makes autonomous work difficult to govern.

CSSL proposes a common account of semantic identity, dependent information, coordinated change, evidence, information flow, and operational history. The central research contribution is their composition: a continuing system should be able to explain what changed, which assumptions supported it, who could authorize it, what was observed, and what remains uncertain. These questions remain distinct even when their answers are recorded together.

This paper develops that account through four illustrative cases: coordinated change, incomplete information, uncertain external effects, and revised evidence. The Sovereign-Closure revision retains open construction, convergence, and verification obligations. Its models and performance targets define requirements for evaluation; a complete implementation and demonstrated conformance remain to be established. The discussion identifies the practical questions such an evaluation would need to answer.

1. Why a history of edits is insufficient

Software work changes more than text. A renamed component may retain its purpose and dependencies. An unchanged component may acquire a different meaning because an interface, policy, environment, or dependency changed around it. A generated artifact can have the same bytes as an earlier artifact while belonging to a different authorized release. Conversely, a small edit can alter the conditions under which a large set of consumers remains valid.

The problem becomes more demanding when autonomous authors work continuously. They receive information through searches, tools, services, and previous work. They produce drafts, request execution, and respond to failures while other authors change the environment. A review that sees only the final textual difference cannot establish which

information was actually available, whether an apparent absence was searched comprehensively, or whether an approval still covers the resulting work.

CSSL therefore treats change as a relationship among an enduring subject, its particular revision, the information supporting the revision, and the authority and evidence associated with its consequences. A record of those relationships can support explanation and checking. Recording them cannot by itself make the information correct or the proposed change safe. This limitation matters throughout the study: a richer ledger improves the questions that can be asked, but the answers still depend on the quality and completeness of its producers.

The collection thesis places this problem within governed autonomy. The architecture study discusses the broader separation between reasoning and authority. Here the concern is the continuity of change across those boundaries. Public review scope.

2. Identity, revision, and meaning

An enduring entity and a particular representation of that entity answer different questions. Identity asks whether this is the continuing subject of earlier work. Revision asks which version is under discussion. Content identity asks whether particular bytes are equal. A name gives people or other components a way to refer to a subject, but a name alone does not establish its continuity.

These distinctions have practical consequences. Renaming a function need not create an unrelated function. Copying its text need not transfer its identity or the approvals attached to it. Deleting a subject and later creating another subject with the same name need not restore the earlier subject. Generated output also requires a lineage account: equal output does not explain which inputs, generator, or environment produced it.

Meaning introduces a further boundary. Different readers depend on different aspects of an artifact. One consumer may depend on a public interface; another may depend on behavior; a formatting tool may depend on exact syntax. A change that is irrelevant to one reader can invalidate another reader's assumptions. CSSL uses the idea of an explicit semantic interpretation to make this dependence discussable. Such an interpretation is itself a contract with limits. It cannot be assumed complete merely because it has a formal name.

The source consequently distinguishes stronger semantic understanding from conservative treatment when understanding is absent. A system must not announce that a change is irrelevant when its analysis does not cover the affected meaning. This is particularly important across language boundaries, generated artifacts, configuration, data, and operational policy. The available interpretation may differ substantially between these subjects.

Semantic continuity is therefore an obligation to preserve and explain relationships through change. It is not a claim that all software meaning can be mechanically captured. Public review scope.

3. Information that actually supported the work

A proposal is conditioned by what its author received. A search result, document, policy instruction, test definition, or environmental observation can influence the work even when it is not mentioned in the final explanation. CSSL asks for a record of governed information delivery so that checking can distinguish the information actually supplied from an author's later account of it.

This has a deliberately bounded meaning. A delivery record can support a claim that particular content was served under particular conditions. It cannot prove that a model understood the content, reasoned correctly from it, or was uninfluenced by its prior training. Nor does it establish that no information arrived through a channel outside the governed environment. The trust boundary must identify which channels are covered.

A coherent view of information is also different from a collection of individually authentic observations. Two authentic responses can describe states that never coexisted. A proposal may depend on both an interface and its consumer, or on a policy and the authority derived from it. If those facts are combined across an incompatible change, checking them separately does not establish that the proposal had a valid basis.

Search illustrates the issue sharply. Approximate retrieval can find useful material without establishing completeness. Returning no relevant result from such a search does not establish that no relevant result exists. An absence claim needs a search whose declared scope and completeness support that claim. Even then, the claim applies to that scope and time. A later addition or a change in the search definition can make the earlier conclusion stale.

This account changes the role of refresh and recovery. When supporting information changes, the proposed system would identify work that needs reconsideration while retaining useful history and unaffected activity. The intended benefit is more precise renewal of evidence. Whether it reduces rework at an acceptable cost remains an empirical question. Public review scope.

4. Coordinated change and accountable authority

Some changes only make sense together. An interface and its consumers may require a coordinated revision. A new policy and the work permitted under it may need a consistent interpretation. Showing only part of such a change can expose an intermediate state that satisfies none of the intended conditions.

CSSL treats coherent visibility as a separate problem from collecting approvals or storing drafts. Preparing a change does not mean that it has become authoritative. Evidence that participants are ready does not mean that ordinary readers may treat a partial result as the completed change.

4.1 An observable criterion for coordinated change

For a declared atomic change, the affected state is the complete set of participating revisions and their dependencies: for example, an interface, its implementation, the applicable policy, and the data representation on which both operate. The relevant readers are ordinary governed reads and the tools, builds, and decisions that rely on those reads. Diagnostic inspection of pending work is a separate observation mode; it cannot supply ordinary decision premises.

The intended invariant is **all-or-none visibility of the declared participant changes at an activation-aware, causally closed view**. A view including the change must include, or be able to obtain, every participating post-state and the dependencies supporting its activation. A view preceding it excludes those post-states. A replica that lacks a required revision may obtain the certified pending state or wait; it may not substitute an incompatible local revision. This is a visibility requirement. It does not itself prove serializability of arbitrary transactions, real deployment success, or soundness of every participating artifact.

Suppose a change declares that interface I1, implementation M1, policy P1, and data representation S1 replace I0, M0, P0, and S0 together. Assume these revisions depend on one another, so no cross-version compatibility has been established.

Reader's observation	Status under that declared change
I0, M0, P0, S0 in a view preceding activation	Permitted as the coherent old view. It cannot be relabeled as current authority after the governing policy changes.
I1, M1, P1, S1 with activation and every required dependency included or obtainable	Permitted as the coherent new view. Actual deployment still needs separate observation.
I1 with M0, or P1 with S0, as an ordinary view of this atomic change	Forbidden: part of the declared change has become visible without the rest.
I1, M1, P1, S1 shown after preparation or application but before activation	Forbidden as ordinary authoritative state; preparation and pending application are coordination evidence.
Pending M1 inspected in a diagnostic view while ordinary reads retain the old view	Permitted as explicitly pending evidence; it cannot establish completion or justify a dependent action.
New activation known, but S1 temporarily unavailable	Wait or recover the required state; do not return S0 alongside the new participants.

This example assumes all four revisions are participants. A separately declared compatible change may leave a component unchanged, and a declared staged change involving an independently controlled external effect has different permitted intermediate observations. Neither may be substituted silently for an atomic declaration. Concurrent work must still satisfy its own conflict and causal conditions; all-or-none visibility does not make conflicting edits compatible.

Interruption before a final decision does not authorize exposing prepared work. After a final decision to commit, the outstanding duty is to recover missing participant state and complete activation. Until that is possible, affected ordinary visibility remains coherent rather than exposing a completed subset. A timeout is not evidence that the decision never happened. Later revocation or policy change is assessed under its current conditions; a historical coherent view cannot renew expired authority.

Authority must also belong to the correct actor and moment. An author can describe an intention and propose a change. That author cannot establish future observations merely by including them in the proposal. An execution service can report what it observed during execution. A review decision can establish that an authorized reviewer approved a particular scope, subject to its conditions. These facts are related without being interchangeable.

The distinction between intent and authority remains central. A desired outcome can justify investigation and planning while still requiring additional authority for a consequential action. An estimate of the work's scope may prove too narrow. Expanding it then requires an appropriate scope decision; the original estimate is not a permanent exemption from review. Draft work and exploratory alternatives can remain useful without acquiring authority over a shared system.

The research question is whether an implementation can enforce these permitted and forbidden observations through concurrency, interruption, and policy change at an acceptable operational cost. Establishing that result requires the activation construction and its conformance evidence discussed in §9. The invariant states the intended evaluation criterion. Public review scope.

5. Evidence supports particular claims

Evidence has a subject, a method, a scope, and conditions of applicability. A test supports claims about the tested cases and environment. A formal argument supports a stated property under its assumptions. A compiler result establishes something about a particular translation. A runtime observation concerns an observed deployment and interval. A human approval expresses an authorized decision. These forms of evidence do not become substitutes for one another because they are collected in one system.

CSSL accordingly resists reducing assurance to a single universal score. Stronger evidence for one claim can remain irrelevant to another. A proof about an abstract model does not establish that a deployed binary implements the model. Reproducible output does not establish that the output is desirable. A monitoring interval without an alert does not establish the absence of a fault outside the monitor's coverage.

The same principle applies to provenance and independence. A signature can identify a signer and bind a statement without making the statement correct. Many agreeing results can share a common implementation, training source, or mistaken assumption. Their number alone does not establish independence. Nested or repeated attestations should not increase the underlying assurance merely by adding layers of presentation.

Currentness introduces another dimension. Evidence may remain a valid historical account while losing its ability to support a present decision. A dependency change, revoked

authority, withdrawn assumption, or newly discovered limitation may require affected decisions to be evaluated again. Erasing the old result would damage the history. Continuing to treat it as current would damage the decision.

This suggests a disciplined form of accountability: preserve what was known and accepted at the time, and separately establish what can support action now. The evidence statement applies this distinction to the publication itself. Public review scope.

6. Effects, privacy, and replay

Permission to perform an action and permission to use particular information are separate constraints. An authorized process can still be prohibited from sending a document to a particular destination or using it for an unrelated purpose. Information restrictions can apply to derived artifacts, search indexes, model inputs, logs, and metadata as well as obvious source files.

CSSL places these concerns alongside execution because data handling can be consequential even when no business transaction occurs. The fact that two tenants possess identical content does not authorize sharing their derived records. The fact that a model can process a document does not establish that the document may be sent to that model. Any release of information needs support from the relevant authority and purpose.

Execution raises a different uncertainty: what did the outside world actually do? A request may be accepted locally while its remote effect remains unknown. A response can disappear after the effect occurred. Repeating the request can be safe for some destinations and duplicative for others. No general promise of exactly one external effect follows merely from keeping a careful local history.

Recovery must therefore depend on the destination's actual guarantees and on available observations. Some actions can be reconciled, some can be safely retried under their conditions, and some require a decision about unresolved consequences. Compensation is another action with its own authority and outcome. It does not erase the original event or necessarily restore the earlier world.

Replay has a narrower purpose. Reconstructing a history or supplying retained observations to an isolated evaluation must not itself repeat live external actions. Moreover, a preserved commitment to historical content does not guarantee that the content remains available. Privacy restrictions, retention limits, and erasure can reduce what can be replayed while leaving some historical evidence intact. A deletion claim must state its covered stores and known destinations; it cannot establish control over every unknown copy or a model's learned parameters. Public review scope.

7. From changed source to observed operation

The source extends semantic continuity into builds, artifacts, releases, deployments, and incidents. This breadth follows from the underlying problem. A reviewable source change does not identify the artifact that reached an environment. A reproducible build does not establish that its release was authorized. A deployment command does not establish the resulting behavior.

The conceptual objective is to preserve the relationships between these stages while keeping their claims distinct. A build record should describe what was built from which declared inputs. An artifact record should identify the resulting subject. A release decision should apply to the intended source and artifacts under current conditions. Operational observations should identify the environment and interval they actually cover. Repair should connect to the incident and the conditions it is intended to restore.

Resource limits belong in this account. Expensive validation can delay work, and an incomplete search or interrupted computation can remain useful. But exhaustion of a budget cannot turn an incomplete result into complete evidence. Work needed for recovery and control also competes for resources; a design that spends everything on new activity can undermine its ability to govern that activity.

Offline work presents a related boundary. A participant may continue drafting and collecting bounded evidence while disconnected. That activity cannot silently acquire current authority over an environment it has not observed. Reconnection requires reconciliation with the current system and its policies. Similarly, rebuilding a derived view should recover the meaning of recorded events, rather than introduce new authority through the view service itself.

These are specification obligations. Their feasibility, responsiveness, and operational cost require implementation and measurement. Public review scope.

8. Four illustrative cases

Four hypothetical cases show what the proposed account of change would need to preserve. Their outcomes are illustrative.

A coordinated change succeeds. A team changes a reporting interface and the consumers that rely on it. The work begins with an identified view of the relevant interface, consumers, policies, and test requirements. Review concerns the proposed scope and its supporting evidence. The intended successful outcome is a coherent revision whose supporting records distinguish the author, review decision, resulting artifact, and later operational observations. A passing test remains evidence for the test's scope. It does not stand in for authorization or prove that the revised service has reached production.

An unsupported absence claim is denied. An assistant searches for callers of a component and receives no result from an approximate search. It proposes removal on the premise that no callers exist. The empty result does not support that premise. The objective can continue through a suitably scoped exact investigation or through a revised change that does not depend on the unsupported claim. The meaningful result is completion of authorized work with an adequate basis, rather than merely recording that an initial proposal was rejected.

An interrupted external action is reconciled. A release action sends an artifact to a destination, but the response is lost. The local process does not know whether publication occurred. Recovery examines the destination's observable state and the conditions of the original action. If reliable evidence establishes completion, the record can be settled accordingly. If the destination supports an appropriate safe retry, a retry may be possible

under current authority. If neither is established, uncertainty remains a real constraint. Replaying the historical record does not issue another publication request.

Later evidence changes a present decision. A dependency analysis supported a release, and a later discovery narrows that analysis's valid scope. The original release record remains part of history. A present deployment decision must consider whether its required assurance is still satisfied. Affected work can receive renewed analysis, additional evidence, or a revised scope. The new evidence does not make the earlier record disappear, and the existence of the earlier record does not settle the new decision.

Together these cases show why continuity requires more than append-only storage. It requires an account of which relationships continue to support which decisions as circumstances change.

9. Threat assumptions, limits, and research questions

The proposal assumes that authors and external services can be mistaken or adversarial. Signed statements require checking; independent observation matters; authority and information flow need enforceable boundaries. Its guarantees would also depend on the trustworthiness of the components that capture information, interpret meaning, enforce policy, and record consequences. Unobserved channels, compromised enforcement, unavailable evidence, and failures beyond the assumed trust model remain material limitations.

The Sovereign-Closure revision makes three construction requirements central to its foundational claims. Each has a distinct obligation and a bounded example against which a proposed construction can be evaluated.

Foundation	Construction requirement and bounded example	Claims that remain conditional
Binding a proposal to its premises and their discharge evidence	In the nonempty-discharge path, the proposal, premise account, and discharge certificate require one another's final content-derived identities. For a proposal requiring one discharged premise, an otherwise valid certificate cannot simply be attached after the proposal's final identity is chosen without revisiting what that identity binds. A qualifying construction must resolve this dependency in a form agreed by producer and verifier.	Canonical encoding and implementation conformance for that path; complete evidence binding for consequential proposals. An empty discharge list avoids this particular cycle without proving the rest of CSSL.
Binding activation to its supporting frontier	The activation identity depends on a frontier that includes that activation. For a two-participant change, having both post-states ready does not by itself establish the activation/frontier relationship. A qualifying construction must distinguish the commitment before activation from the resulting history after it.	Constructible activation records, activation-aware views, and implementation conformance for atomic cross-domain visibility. The all-or-none criterion in §4 remains the requirement to test.

Foundation	Construction requirement and bounded example	Claims that remain conditional
Complete rendering of accepted ordered placements	The retained traversal fragment renders concurrent blocks $A=[a]$ and $B=[b]$ as $[a,b]$, but can omit a valid later insertion $C=[c]$ whose origins are the adjacent a and b . Its checked scope supplies a detector for this omission. A general rendering argument must also cover supported histories beyond the fragment, including moves and tombstones.	Complete deterministic rendering and the general convergence and noninterleaving claims for supported event histories. A projection must reject missing, repeated, or unexpected placements rather than expose a result that omits an accepted placement.

The distinctions among stable identity, content, historical evidence, current authority, and ordinary visibility state the intended architecture. The source's final self-audit classifies the associated foundational statements as requirements without proof. Their stronger construction and conformance claims therefore depend on discharging the obligations above. Formal models and acceptance cases frame that evaluation; neither their presence nor a closure label establishes successful implementation. [Public review scope](#).

Several research questions follow. Can semantic interpretations be precise enough to reduce unnecessary invalidation while remaining conservative about unknown effects? How can the cost of complete dependency checking be bounded without weakening the claim? Which combinations of evidence provide useful assurance for a particular decision, and how should common causes be represented? How can coherent change remain available under realistic interruption and trust failures? What proportion of external uncertainty can be resolved by observation rather than repetition? How should privacy-driven unavailability change the claims a historical record can still support?

The proposed composition is valuable as a research agenda because these questions constrain one another. Faster change is useful only if the resulting evidence remains adequate. More retained information can aid reconstruction while increasing privacy obligations. Finer semantic distinctions can reduce unnecessary rework while making the interpretation itself more demanding to validate. CSSL's conceptual position is that these tradeoffs should be explicit in the system's account of change.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

Foundations. Review the relevant argument and assumptions in this paper, the [evidence guide](#), and the [research review route](#). Specification, reported results and production observations have distinct scope.

Identity and interpretation. Review the relevant argument and assumptions in this paper, the [evidence guide](#), and the [research review route](#). Specification, reported results and production observations have distinct scope.

Information and currentness. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Coordinated work. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Assurance. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Effects and information governance. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Operational continuity. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

IC-POAR: From intended behavior to an accountable running product

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Intent Compiler and Projected Operating Artifact Runtime architecture

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

The Intent Compiler and Projected Operating Artifact Runtime (IC-POAR) investigates a different unit of software authorship: the intended behavior of a product, together with the authority, experience, state, and evidence needed to sustain that behavior. Its proposed output is a running product whose implementation can change while its intended meaning remains identifiable and reviewable. Source code remains important, but a particular codebase is one realization of that meaning.

The architecture addresses a recurring gap in automated development. A system can generate plausible interfaces and services without establishing that they form a complete product, that their controls reach authorized effects, or that the product can continue through interruption and change. IC-POAR makes these relationships part of the intended outcome. It also distinguishes interpreting a request from having authority to enact an interpretation, and distinguishes reconstructing software resources from repeating consequential actions in the outside world.

This paper develops the proposal through definitions, experience and authority boundaries, and four illustrative cases. The founding corpus specifies requirements and a reference trace whose observations and receipts are illustrative; implementation qualification remains to be demonstrated. The central research questions concern faithful interpretation, discovery of missing requirements, affordable verification, human comprehension, and continuity through changes in implementation.

1. The gap between generated parts and a running product

A product request describes an outcome in the user's world. A user may expect work to continue after an application closes, a decision to remain subject to approval, or a report to remain retrievable after a service is replaced. Those expectations concern behavior through time. Producing a screen and a service that look appropriate does not establish that the expectations are satisfied.

Some of the missing work is connective. A visible control needs an operation with an owner, authority, state, failure handling, and an observable result. A background activity needs continuity beyond the interface that started it. An update needs a way to preserve existing records and in-progress work. A polished success display needs evidence that the relevant outcome actually occurred. These are parts of the product rather than optional explanations attached after construction.

IC-POAR proposes that such obligations belong in the maintained account of intended behavior. They should survive changes in framework, service boundaries, storage, and deployment. This does not mean that a product description can contain all knowledge in advance. It means that missing knowledge should remain visible as missing, and that decisions made during implementation should remain connected to the requirements they serve.

The proposal's ambition is therefore broader than producing code from a prompt. It studies whether intended meaning can remain the continuing basis of construction, activation, observation, and change. The [collection thesis](#) explains the surrounding account of governed autonomy, while the [architecture study](#) places this proposal among other work on authority and continuity. [Public review scope](#).

2. Intended behavior, realization, and observation

Three concepts organize the argument. **Intended behavior** describes what a product is required, permitted, or prohibited from doing in a declared context. A **realization** is a particular arrangement of software and resources intended to deliver that behavior. An **observation** records something about what happened in an operating environment.

Their relationships matter because none can simply replace the others. An existing implementation is evidence of how the product currently works; its existence does not settle how the product ought to work. An observation can show that an expectation is unmet without granting permission to lower the expectation. An approved requirement can establish what should happen without proving that the current realization achieves it.

An **operating projection**, in the source's terminology, is a concrete realization for an environment, including the relevant executable parts, state relationships, operational controls, and evidence. The word projection does not imply a disposable imitation. The intended result is a real operating product. It emphasizes that the same product meaning may have more than one acceptable implementation, each subject to its own constraints.

Continuity also requires identity. A product capability can remain the same continuing responsibility when its name or implementation changes. A data record can remain the same subject when its storage representation changes. Equal bytes do not establish that two subjects have the same authority or history. Conversely, changed bytes do not necessarily imply a new product responsibility.

The source's use of “canonical” is deliberately limited. It means that the current account has been resolved under declared rules and authority. It does not mean that the account is complete, infallible, or free of unknowns. This distinction prevents formal organization from becoming a claim of universal understanding. [Public review scope](#).

3. Interpretation does not confer authority

Human intentions reach a development system through mixed material: conversation, written requirements, prototypes, screenshots, existing code, observed behavior, and institutional policy. These materials can be helpful without having the same status. A screenshot may establish an approved arrangement of visible elements while saying little about persistence. Existing code may reveal a compatibility constraint while also containing accidental behavior. A model may propose a reasonable interpretation while lacking authority to decide that it is the intended one.

IC-POAR treats this distinction as a governing boundary. Interpretation produces candidates for acceptance. Authority depends on the subject, scope, context, and decision being made. A person responsible for appearance need not be able to authorize a new external action. An operator's emergency intervention need not amend the product's long-term requirements. A confident model explanation does not settle a conflict among competent decision makers.

History matters here. A later statement can clarify, narrow, supersede, or temporarily override an earlier one. Recency alone cannot determine which relationship applies. Nor can a later correction automatically provide authority for an action already performed. The record needs to preserve what was accepted at the time and what is now understood about that period.

The practical challenge is to keep this account useful to people. Ordinary technical choices should be resolved within existing authority. Decisions that materially alter the outcome, data use, risk, experience, or permitted consequences need appropriate human judgment. Determining that boundary reliably is itself a research problem. The proposal does not assume that asking more questions automatically improves governance; unnecessary questions can obscure the consequential ones. *Public review scope.*

4. Discovering requirements without inventing intent

Product requests leave many implications unstated. If work must survive an interface closing, its continuation cannot depend solely on that interface remaining alive. If an action requires approval, the approved action must remain identifiable when it is carried out. If users must be able to recover their records after an update, continuity of those records becomes part of the update's obligations.

IC-POAR calls the accounting for explicit and derived requirements **closure**. An obligation is a requirement that must receive an adequate disposition before the relevant outcome can be accepted. The concept is useful because it directs attention to missing connections and consequences, rather than treating a list of completed components as sufficient.

Derivation nevertheless has a danger: a familiar implementation pattern can be mistaken for a logical necessity. A promise of durable work may require durable responsibility and recovery, but it does not by itself select a particular vendor or framework. A successful pattern in one environment may be unsuitable in another. The source distinguishes requirements from preferences, assumptions, and proposed hypotheses to keep these differences reviewable.

The scope of completeness must remain bounded. No finite catalogue can be presumed to discover every requirement in every domain. A system may account completely for the obligations it has represented while still missing an important implication of the original intention. The research agenda therefore includes independent evaluation of missed requirements and unsound derivations, not only checking whether a generated plan satisfies its own description.

This is one of the proposal's central tensions. Greater automation depends on deriving more of the necessary work. Greater accountability depends on refusing to silently turn uncertain interpretation into authoritative meaning. A practical system must improve both together. Public review scope.

5. Experience is part of behavior

The experience of a product includes what people can perceive, understand, navigate, operate, and recover from. It includes focus, keyboard access, responsive behavior, content, motion, status, and the availability of actions under different conditions. A screenshot covers only part of this subject.

IC-POAR's **experiential contract** expresses the approved aspects of that experience and connects them to product behavior. A control has a purpose beyond responding to a click. A completion message makes a claim about an outcome. A loading state, interrupted connection, empty workspace, expired session, or unavailable action affects what the user can do next. These relationships should remain coherent when the implementation changes.

Prototypes create a particular interpretation problem. A prototype may use sample data, timers, local state, or scripted feedback to demonstrate an interaction. Approval of the experience does not automatically approve those mechanisms as the production behavior. At the same time, replacing them with real operations must preserve the aspects of the experience that were actually approved. Functional completeness and experiential fidelity constrain the same product.

Accessibility makes the limitation of visual similarity especially clear. Two screens can look alike while differing in keyboard operation, reading order, focus recovery, or the information available to assistive technology. Conversely, two platform-specific controls may look different while providing an appropriately equivalent action. Acceptable variation depends on the declared experience and the people using it.

Status also has meaning. An estimate can be useful if presented as an estimate. A pending request can be useful if its completion remains distinguishable. When observations become stale, continued visual motion should not imply that fresh progress has been observed. The source seeks evidence-connected presentation while leaving the effectiveness of its proposed verification approach to be established through implementation and evaluation. Public review scope.

6. Continuing through implementation change

If intended behavior is to outlast a particular implementation, change must preserve more than a feature list. Existing records, authority, in-progress activity, and previous consequential actions must remain intelligible. A replacement service that starts correctly can still be incompatible with data written by its predecessor. Routing users back to an older version can still be unsafe after a newer version has changed that data.

The architecture therefore treats continuity as a property to examine during change. State has a continuing semantic subject; storage is one binding of that subject. A migration must account for what it preserves, transforms, or loses. Information loss and irreversible change require explicit treatment. Operational rollback, data repair, and a decision to reverse intended behavior are different actions, even when an incident makes them closely related.

External effects impose an additional boundary. Recreating an owned software resource is different from repeating an email, payment, deletion, or physical command. A historical record may show that an action was attempted while leaving its result uncertain. Restoring a software configuration does not restore the outside world, and absence from a local view does not prove that the action never happened.

Recovery consequently depends on what can be observed and on the actual guarantees of the external destination. A safely repeatable operation may permit another attempt under its conditions. An uncertain irreversible action may require further reconciliation or a decision that the system cannot make alone. Compensation is a new action with its own consequences, rather than erasure of the earlier one.

The proposal also considers partial connectivity. Disconnected work may continue within a bounded authority and information context. It cannot receive immediate revocations or assume that remote conditions stayed unchanged. The eventual reconciliation must respect those limitations. These are requirements of the proposed architecture, not operational guarantees demonstrated by this paper. [Public review scope](#).

7. Collaboration and verification

Parallel contributors can produce changes that are individually sensible and jointly incompatible. Their files may not overlap while their assumptions about ownership, authority, timing, or permitted behavior conflict. Agreement on text is therefore insufficient to establish agreement on a product's meaning.

IC-POAR studies contributions as proposed changes to that meaning. A contribution has an intended objective and a basis that can become stale. Updating it for a changed environment should preserve the objective or make the need for reinterpretation explicit. Historical reputation can inform review effort, but it cannot replace evidence or confer authority over a different subject. Useful incomplete work may remain useful without being treated as a completed capability.

Verification then asks what evidence is adequate for the particular obligation. Tests, bounded analysis, mathematical proofs, accessibility checks, operational observations, and human acceptance support different claims. A strong result for one component does not

make every dependency equally strong. A signed receipt identifies a statement and its provenance; it does not make the statement universally correct.

The proposed certification claim is correspondingly conditional. It concerns stated inputs, assumptions, methods, bounds, and an environment. Evidence can lose present applicability when one of those conditions changes while remaining part of the historical record. Reusing it requires support for continued applicability, rather than matching the name of an earlier test.

Verification must also examine prohibited behavior. A test run that never encountered an unauthorized action does not necessarily establish that the action is unreachable. The strength of an absence claim depends on the space examined and the method used. Similarly, a system that passes narrow tests may still exploit gaps in what those tests express. Independent review, counterexamples, and evaluation against intended meaning remain essential research concerns. [Public review scope](#).

8. Four illustrative cases

Four hypothetical cases illustrate the relationship between intended behavior and its realization. They are examples for analysis rather than reports of executed systems.

A continuing research task. A user requests a research task that must continue after the interface closes and remain available when the user returns. A complete realization would preserve the work's identity, authorized scope, results, and recovery behavior beyond the initiating interface. Returning to an attractive dashboard alone would not establish success. The relevant success case is continued authorized work followed by an accessible result whose displayed status has an adequate basis. The example leaves the implementation choice open.

A useful interpretation is denied authority. A proposed enhancement would automatically send the completed research result to a wider audience. The enhancement may appear convenient and technically feasible, but it changes the consequence of the user's request. The system can retain the proposal for review without treating it as permission to send. Work already authorized can continue, and the result can be made available through its approved route. The boundary concerns the additional action, rather than the value of the suggestion.

A replacement preserves the user's records. A product moves to a different implementation while users have existing records and ongoing activity. The intended outcome is continuity of the subjects and their required behavior. A successful installation is insufficient if records disappear, authority changes silently, or earlier actions are repeated. If the transition encounters a recoverable inconsistency, the system needs an authorized continuation that respects the actual state. Whether that continuation is a compatible rollback or forward repair depends on what has already changed.

An apparent success lacks evidence. A prototype shows a completed operation after a short animation. During construction, the animation is replaced by an actual service call, but the interface still announces completion before the outcome is observed. Visual similarity has preserved the wrong aspect of the prototype. The appropriate correction connects

presentation to the intended operation's status and gives the user a usable experience while completion remains pending or uncertain. This illustrates why approved appearance and approved behavior must be evaluated together.

9. Evidence boundaries and open questions

The founding corpus is an architectural specification and research programme. Its reference trace follows an intended product through meaning, construction, evidence, and operation. The example digests, signatures, observations, and receipts illustrate the proposed form; they are not records of an executed deployment. Proposed integration with another system similarly requires evidence about that system and the resulting operation.

The corpus classifies design problems as buildable, research-assisted, or open in their universal form. These classifications concern the availability of possible mechanisms. The research thresholds are proposed evaluation criteria. Compiler performance, independent conformance, live activation, migration behavior, and human comprehension still need to be measured in an identified implementation.

The threat assumptions are substantial. Sources and models can be malicious; generators and verifiers can be defective; authority roots can be compromised; observations can be forged or incomplete; external providers can change behavior. A design that separates these roles still needs evidence that its boundaries are enforced. Shared failure causes can also weaken the apparent independence of multiple checks.

Several questions are especially consequential. Can people approve the meaningful consequences of a change more accurately than they can review conventional specifications and source? Can a system preserve materially different interpretations without overwhelming either the user or the computation? Can it discover missing requirements without treating preferred architecture as necessity? Can verification remain affordable as behavior spans interfaces, services, state, and external effects? Can localized updates reuse unaffected evidence without overlooking hidden dependencies?

The proposal draws on compiler verification, model-driven engineering, capability security, incremental computation, provenance, declarative control, and accessibility practice. Its contribution is the proposed relationship among lasting product meaning, authorized change, complete operation, and evidence of consequence. Bounded implementations and independent evaluation would test whether that relationship improves outcomes and makes consequential decisions easier to understand. The evidence guide describes the standards appropriate to those claims.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

Foundations. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Semantic model. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Authority and decisions. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Requirements and construction. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Experience. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Continuity. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Contributions and assurance. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

AD8: Accountable change across software, action, and learning

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the AD8 specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

An autonomous system can change the software it runs, take actions through that software, and use the resulting experience to change its future behavior. Each activity has a different evidential and authorization burden. Combining them without preserving those differences allows a successful build to stand in for a successful deployment, an apparently useful action to stand in for an authorized action, or accumulated usage to stand in for permission to train. AD8 investigates a common account of identity, authority, evidence, effects, and continuity across these activities. Its contribution is an architectural proposal for keeping their responsibilities connected without allowing one responsibility to confer the powers of another.

This study examines intended meaning and observed behavior, the continuing validity of decisions, agreement among generated artifacts, and the authorities involved in learning and model release. Four illustrative cases connect these questions to ordinary work and recovery. The available evidence includes reported reference checks and bounded formal results. Complete conformance remains unestablished, and a construction affecting a foundational evidence relationship remains unresolved. The paper identifies what further evaluation would need to demonstrate.

1. The problem of connected responsibility

Automation is often organized around individual transactions: accept a request, produce a result, and report success. Long-lived autonomous work is harder to describe this way. A request can cause a new artifact to be generated; that artifact can depend on other software; the software can act in an external system; the resulting observations can influence later plans or become proposed training material. Responsibility follows this chain even when the organizations, technologies, and times involved change.

A plausible result at one point does not settle the questions at the next. A generated document might satisfy a formatting test while containing an unsupported factual claim. A tested program might be unsuitable for the environment in which it is deployed. A payment request might be authorized but have an uncertain outcome after a connection loss. A useful transcript might contain material whose owner permitted operational use but prohibited shared learning. These are different conditions, requiring different evidence and different authority.

AD8's master specification proposes that those distinctions remain part of the system's enduring account of its work. The central question is not simply whether a task can be completed. It is whether a particular change or action can be justified by the intended meaning, current authority, applicable evidence, and known external conditions that bear on it. [Public review scope](#).

The proposal spans software realization, concurrent work, external effects, stochastic generation, learning, model release, and recovery. Evaluating it requires both evidence for those responsibilities and evidence that their boundaries survive composition. The [architectural overview](#) compares AD8 with related works, each of which retains its own scope and requirements.

2. Identity, meaning, and authority

The first conceptual distinction concerns what a thing is, which version is being discussed, what information supports it, and what it is presently allowed to do. These questions are related, but a single label cannot safely answer them all.

A continuing business object can change while remaining the same object for organizational purposes. A particular version can remain available as historical evidence after it is no longer eligible for current use. A signature can establish a relationship to a signing identity without establishing that the signer had the relevant power, that the intended audience matches, or that the permission remains valid. AD8 consequently treats identity and current authority as separate subjects of reasoning. [Public review scope](#).

An intention is also distinct from evidence about the world. An organization may intend to restrict an action to a class of recipients; observations may show that a candidate action includes someone outside that class. The observation does not silently amend the intention. Conversely, the existence of an intention does not establish that the world already conforms to it. This separation enables disagreement and uncertainty to remain visible instead of being resolved by whichever representation is most convenient to the next component.

Authority is contextual. It concerns an actor, a purpose, an affected resource or party, and relevant conditions. Delegation is meaningful only within those limits. Passing a request through another component should not enlarge it, and a technical ability to perform an operation is not sufficient justification for doing so. Recovery likewise cannot create a permission merely because the ordinary permission service is unavailable.

The practical consequence is that historical validity and current usability need separate answers. A reviewer may legitimately inspect a withdrawn decision in order to understand prior events. That historical access does not make the decision suitable for a new action. This distinction is particularly important when a task lasts longer than the permissions, environmental assumptions, or evidence on which its initial plan depended.

3. From intended behavior to an operating system

AD8 treats software realization as an accountable interpretation of intended behavior. The intended result includes more than the most visible feature. It can include state ownership, access restrictions, data preservation, recovery, accessibility, operational behavior, and the evidence needed to determine whether the result is acceptable. A plan that accounts for the happy path while omitting these obligations is incomplete in the proposal's terms. **Public review scope.**

This conception of completeness does not imply that a compiler can discover every relevant fact or settle every human judgment. Some obligations depend on research, legal interpretation, or a decision among materially different alternatives. Those dependencies must remain identifiable. A favorable score on convenient criteria cannot compensate for a prohibited action or an unresolved requirement that is essential to the intended result.

The proposal also separates the artifact from its operation. Preserving the bytes of a generated artifact establishes which artifact is under discussion. It does not establish its semantics, suitability, or successful activation. Testing contributes evidence within its stated scope. Deployment introduces additional questions about the environment, state transitions, compatibility, and observed behavior. A partially activated system cannot be described as fully operating simply because some of its components are healthy.

The same distinction governs repair. A system may be able to reconstruct a resource that it owns and whose intended state is established. It cannot infer that every absent external effect should be recreated. An absent local record of a message does not establish that the message was never sent. Nor can repair overwrite a person's work merely because that work differs from a desired configuration.

Experience is part of this account. If a promised behavior requires a person to perceive a status, use a control, recover focus, or understand a necessary approval, then producing backend machinery alone does not satisfy the promise. AD8 includes experiential obligations while retaining a separate authority boundary: an interface can communicate an authorization decision, but its appearance cannot establish the underlying permission.

4. Concurrent work and the validity of premises

An autonomous worker reasons from a view of the world. Other workers may change that world while the reasoning continues. Correctness therefore depends on more than whether a view was once accurate. It also depends on whether the worker consulted the information necessary for its decision and whether relevant changes have occurred since then. **Public review scope.**

Consider a task that prepares a change for a group of accounts. Reading every account already known to the worker does not necessarily establish that the group is complete. A newly eligible account or a change in the rule defining the group can matter just as much as an edit to a known account. Adequacy of the premises and freshness of those premises are distinct questions.

AD8's proposal gives both questions architectural standing. The conceptual requirement is that a consequential change remain related to a sufficiently complete and sufficiently

current basis. When that relationship no longer holds, the system must reconsider affected work. Merely recording that a worker read something earlier cannot justify a later decision under changed conditions.

This does not require every concurrent change to invalidate all work. Some changes are independent under a declared interpretation. The research difficulty lies in establishing that independence rather than assuming it from superficial separation. Two edits in different places can still conflict through a common business rule; two edits to the same object may concern independent properties. Any stronger claim about compatible concurrency depends on the semantics and evidence available for that domain.

The proposal thus treats collaboration as a problem of preserving justified decisions over time. Scheduling mechanisms may reduce wasted work, but scheduling priority does not create authority or settle a semantic conflict. The CSSL study develops the related distinction between changing stored content and changing justified semantic state.

5. Effects, observation, and uncertain outcomes

An effect is a consequence that matters beyond an isolated computation. It can involve data, communication, money, physical activity, or a change in another service. AD8 argues that effects should be considered both before execution and through observation afterward. Intended scope, authorized scope, and observed consequences answer different questions. Public review scope.

Different powers cannot always be reduced to one scale. Permission to read a file does not imply permission to send its contents elsewhere. Permission to spend within a budget does not imply permission to contact any recipient. A system that reports only a broad confidence or risk score can conceal those distinctions. AD8 instead investigates a structured account in which the relevant dimensions remain separately meaningful.

Attribution is equally important. An action can pass through several components before reaching an external service. Its consequences do not lose relevance because a supporting component produced them. Operational measurements may themselves become significant if they influence later decisions, expose information, or alter consequential state. The classification must follow actual use and consequence, rather than the component's convenient description of itself.

Observation also has limits. If a connection ends after an external request may have been dispatched, the system cannot conclude either that the action succeeded or that nothing happened. Repeating it may duplicate a payment, message, reservation, or physical instruction. Recovery must establish what happened and which further actions remain authorized. A compensating action creates a new consequence; it does not make the original event disappear.

This is a demanding form of continuity. The objective remains relevant after a failed attempt, but continuing requires a valid account of current conditions. The system may resume, refresh evidence, choose another eligible route, or reconcile an external result. It must not confuse persistence with permission to repeat an action whose outcome is unknown.

6. Generated agreement and bounded assurance

Stochastic systems can produce several plausible artifacts for the same intended task. Their agreement is informative only in relation to what was held constant, what was varied, how results were compared, and which failures remained visible. AD8's certification proposal is motivated by the risk of converting repeated plausibility into an unwarranted claim of correctness. Public review scope.

Two generated results can appear equivalent under the observations available to an evaluator while differing in an untested condition. Evidence that has not distinguished them is weaker than a proof that they are equivalent within a defined model. A newly discovered distinguishing case can therefore matter more than a large collection of earlier similarities. The public claim must remain bounded by the interpretation and instruments that supported it.

Agreement also differs from compliance. Many candidates may converge on the same prohibited behavior. Conversely, several distinct realizations may all meet the requirements. Combining agreement and compliance into an undifferentiated success measure would obscure both possibilities. The proposal keeps their evidential roles distinct and requires variation across conditions to remain visible rather than disappearing into an average.

This is not a claim that deterministic checking eliminates uncertainty. A checker has a scope, assumptions, and an implementation of its own. The choice of observations may miss a relevant consequence; a test population may inadequately represent future use. The research question is how to state the resulting guarantee precisely enough that another party can tell where it applies and where further evidence is needed.

The comparison with SEPION concerns assurance for generated artifacts. It identifies related evaluation questions; transferring a guarantee between realizations would require an explicit argument about the relevant models and assumptions.

7. Learning, model release, and continuing eligibility

Operational experience can support improvement, but use of a system is not itself permission to contribute to shared learning. AD8 distinguishes ordinary use, private memory, organization-specific adaptation, and shared model development. Each raises different questions of purpose, ownership, privacy, and future influence. Public review scope.

The proposal treats an experience as more than its successful ending. Failed branches, corrective actions, uncertainty, dissent, and missing observations can all change what may reasonably be learned from it. Removing those elements can turn a conditional success into an apparently general rule. An account of experience should preserve the limits of its causal interpretation, including whether the claimed outcome was independently observed.

Even eligible material need not belong in shared model parameters. Some information changes frequently, applies only to a particular organization, remains private, or requires prompt withdrawal. Its appropriate use may remain external to the model. More persistent learning raises a stronger burden because the resulting influence can be harder to isolate

and remove. This is a question of the proposed destination and future use, rather than an assumption that more collected material is always better.

AD8 further distinguishes producing a model candidate, evaluating it, making it eligible for release, deploying it, and selecting it for a particular request. A candidate's capability gain does not settle privacy, authority, security, or lineage concerns. The ability to train a model cannot confer the right to approve its own use. Nor does availability of a model imply that it is suitable for every organizational or jurisdictional context.

Continuity of model assets introduces another separation. Preserving or reconstructing model bytes can be necessary for recovery, but possession does not itself authorize deployment or behavior. Similarly, withdrawing an eligible contribution does not prove that every past influence has been erased from a trained model. Future-use restrictions, replacement, validated remediation, and residual influence require distinct statements. Claims of complete removal must be supported by the evidence actually available.

8. Four illustrative cases

Four hypothetical cases illustrate the responsibilities AD8 seeks to connect. Their outcomes are examples for analysis rather than recorded executions.

8.1 An authorized service change

An organization requests that customers receive a clearer appointment confirmation. The intended outcome includes the message's meaning, permitted recipients, accessibility, and preservation of existing appointments. A proposed realization accounts for these obligations, and review evidence addresses the behavior in the intended environment. The organization authorizes the change within a defined scope.

Success in this illustration requires the relevant service to operate as intended and the result to be observed. A stored artifact or an accepted deployment request is an intermediate fact. If only some routes use the new behavior, the account remains partial. The example demonstrates why semantic intent, artifact identity, operating state, and observed experience need connected but distinct descriptions.

8.2 A useful proposal outside its authority

A worker preparing the confirmation improvement proposes adding a promotional message to every customer record. The suggestion might be commercially attractive, and its implementation might pass technical checks. Neither fact establishes permission to use the broader recipient group or to change the message's purpose.

The expansion is denied within the existing authority. Work on the authorized confirmation improvement can continue. If the organization later chooses the broader change, that choice needs its own adequate basis and authorization. The denial concerns the particular proposed expansion; it does not redefine the original objective as failed or require abandoning work that remains achievable within scope.

8.3 A payment response is lost

An authorized payment is submitted, but the connection ends before its result is observed. The worker retains the objective of settling the invoice. It does not infer from the missing response that the payment never occurred.

An authorized inquiry establishes that the external service completed the payment. The local account is reconciled to that observed result, and the invoice can be reported as settled. If the inquiry had instead established that no payment occurred, a permitted continuation could be evaluated. If the result remained unresolved, the account would retain that uncertainty. The example illustrates recovery through knowledge of the external state rather than through blind repetition.

8.4 A learning contribution is withdrawn

A participant explicitly allows a bounded body of experience to be considered for a learning purpose. Later, the relevant permission is withdrawn. The system must determine which proposed or active uses depend on that contribution and prevent uses that no longer satisfy the applicable conditions.

The claim that can be made depends on the affected destination. Removing an external reference, retiring a candidate, and addressing influence in an already trained model are different accomplishments. The system records what it has actually restricted or remediated and which residual questions remain. Recovery aims to restore an eligible service or model route without silently preserving a use whose justification has been withdrawn.

9. Evidence boundary and open questions

AD8 2.0.0 defines a proposed architecture and its conformance requirements. Its disclosed evidence includes reference encoding checks, generated-representation checks, and three bounded formal models executed in TLC. Authored traces remain illustrations. The results below establish their stated finite propositions and are attributed to the recorded source runs. Public review scope.

9.1 The three finite model results

The model scope guide and the recorded 22 September 2026 verification identify TLC 2.19, the selected model and configuration revisions, and four deliberately faulty controls. The table reports those records, rather than inferring a result from the presence of formal source. Every successful base model has four distinct reachable states and complete search depth three. Their small size is part of the scope, not evidence of a complete AD8 implementation.

Model and proposition	Domain and premises	Reported checking result and control	Claim supported
Approval payload binding	One approval, two possible presented payload identities, and one consumption; execution must consume the approved payload	Base model: no error, 10 generated states and four distinct states. Removing the payload-equality guard produces an invariant counterexample.	The executed payload equals the approved payload in this model. A later presented input cannot redefine what was consumed. This does not establish the real signing, approval, or execution path.
Receipt coverage and freshness	Three Boolean properties represent coverage, freshness, and present eligibility; invalidation must remove eligibility	Base model: no error, 10 generated states and four distinct states. Retaining eligibility through invalidation produces a coverage/freshness counterexample.	Current eligibility implies covered, fresh supporting reads in this model. Invalidation affects present reliance; it does not erase historical publication. Completeness of real observations is assumed rather than proved.
One-time initialization and progress	Four abstract phases and one initialization exception; ordinary authority follows initialization. Eventual resolution assumes weak fairness of preparation and finalization-or-abort. Terminal phases permit stuttering.	Base model: no invariant or temporal error, six generated states and four distinct states. Retaining the exception produces an invariant counterexample. Removing fairness produces a temporal counterexample.	The exception cannot remain usable after finalization; initialization precedes ordinary authority; eventual resolution holds under the declared fairness assumption. Real people, services, custody, and recovery are not shown to satisfy that assumption.

The recorded run used TLC 2.19 from the publicly distributed tla2tools 1.7.4 archive (SHA-256 936a262061c914694dfd669a543be24573c45d5aa0ff20a8b96b23d01e050e88). Exact native model/configuration and private result-record identities remain in the private audit. The table preserves the dated propositions, bounds and recorded outcomes; public reproductions below use separate teaching examples.

The deliberate faults are successful controls because the checker produces the relevant counterexamples. A parsing failure or timeout would not supply that evidence. The remaining TLA+, Alloy, Lean, and SMT artifacts are candidate studies without qualification for the broad claims previously attached to them. Neither these three results nor the source's aggregate status counts establish its 215 system-level assurance claims. Concrete behavior still requires a faithful implementation, a stated fault model, and observations at the real decision and effect boundary.

9.2 Foundational dependencies and their consequences

Two disclosed dependencies restrict the transition from reference arithmetic to object conformance.

Foundation	Missing step and why it matters	Dependent claims	What remains supportable
Reciprocal binding between a candidate and its required evidence	The proposed candidate identity depends on a commitment to required evidence, while that evidence identifies the same candidate. The source has not supplied an acyclic construction for a nonempty evidence set. Computing each commitment correctly does not determine how to construct both mutually dependent identities.	Conformance of nonempty evidence-bearing candidates and the authorization or transitions relying on their binding	Bounded encoding and evidence-tree arithmetic; the requirement to bind evidence to the exact candidate. Neither establishes a constructible complete signed object.
Agreement between primitive meanings and generated carriers	The source records a 128-bit decimal description paired with 64-bit carriers and differences in some domain and identifier ranges. A generated representation has not thereby preserved the complete declared value domain.	Full primitive closure and semantic equivalence of every generated representation	Checks over their explicitly supported values and representations. A result for that subset cannot qualify values outside it.

For the first dependency, consider an abstract artifact A whose identity must bind its required review evidence E. If E must already name A's final identity, neither final commitment follows simply from having the other's placeholder. Leaving out the required evidence would change the requirement. For the second, a value accepted in a wider declared range can be unrepresentable in a narrower carrier; successful small-value round trips cannot settle that case. These examples explain construction and compatibility gaps, not newly observed product incidents.

The source also distinguishes historical integrity fixtures from current conformance. Older signature fixtures bind opaque descriptive bytes, and a retained evidence-tree fixture uses legacy text identifiers. Their integrity or arithmetic can still be checked within that historical scope. They do not qualify the current semantic producer, current authority, or typed-object binding. Resolution requires a defined construction and agreement among its producers and verifiers, followed by checks of the affected objects. The three finite models above do not supply that missing construction.

Several research questions follow. How can an observation model be shown to cover the consequences that matter in a new domain? Which concurrent changes can be proved irrelevant to an ongoing decision? How can independent evaluation avoid common sources of error or institutional capture? What evidence justifies a claim about residual influence after learning permissions change? What recovery guarantees remain achievable when the external service or authority required for reconciliation disappears?

There is also an integration question: the more responsibilities a common architecture connects, the greater the need to show that their boundaries survive composition. Local checks can all pass while an end-to-end path silently bypasses an intended owner.

Demonstrating the intended separation therefore requires observed system behavior across success, denial, interruption, and recovery, together with explicit assumptions about the people and services outside the system.

AD8 contributes an account of responsibility across software change, action, learning, and time. Its evaluation would test whether those connections make continuing work more accountable without losing useful progress. The evidence guide, institutional thesis, and glossary provide the wider context for that investigation.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

Scope: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Identity and authority: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Software realization: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Concurrent work: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Effects and recovery: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Generated artifacts: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Learning and models: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

Evidence limits: Review the relevant argument and assumptions in this paper, the evidence guide, and the [research review route](#). Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Freehold: Autonomous institutions under bounded stewardship

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Freehold specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Freehold investigates a demanding form of autonomy: an institution that develops its own internal work while remaining subject to external limits on resources, access, and effects. Its institutional design requires citizens to persist across individual model sessions; its projects, decisions, expenditures, and histories belong to an enduring institution. A human may establish its founding conditions and exercise stewardship over the environment. The human does not subsequently assign its citizens missions or rewrite their decisions. This distinction makes Freehold a study of bounded institutional agency rather than a general description of an assistant completing delegated tasks.

The specification connects this autonomy question to a second problem: how an institution can become intelligible without observation becoming another form of control. The Observatory presents activity and evidence; the World Atlas makes selected institutional histories publicly inspectable; the Exchange describes institution-originated economic and diplomatic relations; stewardship provides custody without an unrestricted channel for directing internal intent. Across these surfaces, a persuasive account of activity must remain distinguishable from an authorized action and from a confirmed external result.

This paper develops those commitments through four hypothetical scenarios and an examination of independence, interpretation, evaluation, and accountability. The source works specify intended behavior; the actual engine and host supplied an implemented research subject in the examination I directed on 10 October 2026 with AI assistance. Demonstrating sustained internally originated work, economic viability, and effective stewardship remains an empirical task. The paper identifies the observations that would make those claims assessable.

1. The problem of a continuing institution

An ordinary delegated task starts with a request. A user describes an outcome, grants whatever authority is necessary, and expects a system to pursue that outcome. Persistence, recovery, and tool use may make the task lengthy, but the origin of the objective remains comparatively clear. Freehold asks a different question: what would it mean for an artificial institution to originate and maintain its own internal agenda within conditions established by someone else?

That question cannot be answered by leaving a model running. Continued generation does not establish a continuing institution. An institution needs a history against which decisions

can be understood, resources against which commitments can be assessed, and an account of which activities belong to which participants. It also needs boundaries that survive a change of model, an interruption, a change of external access, or the abandonment of a project. Otherwise apparent continuity may be only a narrative reconstructed after the fact.

Freehold therefore treats autonomy as a relationship among internal initiative, external authority, and retained evidence. The institution may consider a new venture. That consideration does not establish permission to spend, publish, contact someone, or change the environment in which it operates. Conversely, a restriction on external access need not determine which questions its citizens may consider internally. These are connected concerns with different owners.

The word *institution* is used functionally here. It refers to persistent participants, shared history, internal projects, and bounded commitments. Terms such as *citizen*, *estate*, and *treaty* are vocabulary within the specification. They do not establish legal citizenship, personhood, sovereignty, contractual capacity, or rights over another party. Any real external obligation would still depend on applicable human authority and the conditions of the external service or relationship.

The central contribution is thus a specification of constraints under which internally originated work could remain recognizable as such. The architecture does not establish that initiative will be useful, independent in every causal sense, or economically productive. It makes those questions more precise by asking what is allowed to shape an institution, what is allowed to act for it, and what evidence may support claims about its history. **Public review scope**

2. A vocabulary for bounded institutional agency

The *estate* is the institution's bounded operating territory. In conceptual terms, it identifies the resources and environment within which the institution may act. The *founding conditions* describe the initial purpose, limits, and participants. A *citizen* is a continuing participant whose history extends beyond one response or one model session. A *mission* is internally originated work; a *venture* is a more durable project with a continuing account of purpose, resources, and outcomes.

These definitions matter because they prevent several tempting substitutions. A model instance is not automatically a citizen. A list of generated tasks is not automatically a mission history. A proposed business idea is not automatically a venture with actual revenue. A detailed biography is not evidence that the attributed actions occurred. Each term acquires meaning through continuity and attribution, not through evocative language alone.

The *treasury* represents accountable resources. It supports a distinction between money available, money committed, money spent, and money actually received. The *gates* represent governed external access. Their conceptual purpose is to make the availability of an external action explicit and bounded. A gate is not a reward for eloquent reasoning, and an institution's desire to use a resource does not open it.

Stewardship names the human role in maintaining the conditions of institutional operation. A steward may establish an institution, supply resources, decide whether particular external access is available, inspect evidence, and pause or end operation. These are consequential powers. Freehold nevertheless distinguishes them from directing a citizen's next thought, assigning a mission after founding, choosing an internal vote, or writing the institution's strategy.

Finally, a *chronicle* is an account derived from institutional history, while *public proof* is a selected, privacy-conscious presentation of supporting records. Neither term means that every statement is beyond dispute. Evidence has scope, and a history may establish that a proposal was made without establishing that the proposal was sound. Public proof may support the existence of a recorded event while leaving private information unavailable to an outside reader.

The vocabulary makes the institution inspectable without erasing uncertainty. It also supports comparison with the broader [governed autonomy thesis](#) and [architecture study](#): the ability to generate a course of action, the authority to perform it, and the evidence that it occurred remain separate questions. Shared terms are explained in the glossary.

3. Founding, continuity, and the limits of independence

Founding is an intervention. Initial participants, stated purpose, available resources, and permitted external relations shape what an institution can become. Freehold does not eliminate that influence. Its autonomy claim concerns the distinction between bounded founding conditions and continuing human direction of internal work.

After founding, the specification reserves mission formation, debate, project proposals, and other internal initiatives to the institution. A human's power over environmental conditions remains real, but it is not represented as a citizen's preference. If access to an external resource is withdrawn, the institution encounters a changed condition. The withdrawal should not become a fabricated internal decision that the institution never made. This attribution boundary matters for both research and accountability.

Persistent citizens introduce a related issue. A participant's competence can change as evidence accumulates. A founding description is an initial condition, not proof of skill and not a guarantee of permanent specialization. The source proposes that the institution's account of competence should follow recorded contributions and outcomes. Conceptually, this permits development without treating a self-description as an achievement.

Growth also creates obligations. Adding a participant consumes resources and changes the institution's internal relationships. Retiring or reorganizing participants should not erase responsibility for prior work. Continuity is therefore more demanding than maintaining a name: the history of contributions, commitments, and unresolved matters must remain intelligible when the population changes.

There is an unavoidable philosophical limit to the independence claim. Resources, access, selection of founding conditions, and the availability of information can all affect later choices. A system may exclude direct instructions while still being strongly influenced by its environment. Freehold's narrower, testable commitment is to distinguish permitted

environmental influence from a hidden continuation of direct task assignment. It does not establish causal independence from the people who created and maintain the environment.

An evaluation should therefore ask both whether prohibited direction occurred and whether the remaining channels create systematic incentives or dependencies. These are separate tests. A resource reduction can legitimately change what is feasible, and evidence of that effect belongs in an assessment of dependence. A hidden instruction to select a particular venture instead changes the claimed origin of institutional work. Relabeling the instruction as a resource note or market signal does not make it permitted.

The narrower autonomy test follows the meaning and route of the influence. It asks whether an external act assigned, selected, or rewrote an internal mission, decision, or strategy without the institution's own originating decision. It also asks whether observational preferences or rejected commands were later fed into internal work as incentives or instructions. The absence of a visible conversation box answers only a small part of that question. It says little about other channels, measurement coverage, or how the institution interprets the world it is permitted to see. Public review scope

4. Internal initiative and external authority

Freehold's proposed institution may have reasons for acting that are its own within the specification's operational meaning. It may still lack authority to act externally. This is a necessary condition of bounded autonomy: initiative cannot serve as its own permission.

A reviewer needs separate answers to four questions: what opportunity was understood, why it was pursued, what authority permitted the attempt, and what actually happened. These distinctions apply to research, publication, financial activity, and relations with other institutions.

Authority also has scope. Access to public information does not imply permission to use an authenticated account. Permission to prepare material does not imply permission to publish it. A financial allocation does not imply permission to use every available external route. A reliable history of earlier activity does not provide a permanent grant for later activity under changed conditions.

The institution's record should preserve this difference between desire and permission. A declined external action need not mean that the underlying internal question has ceased to matter. The institution may continue research within its remaining authority, revise a project, or leave an external dependency unresolved. It cannot transform the absence of permission into permission by redescribing the same action.

At the same time, temporary execution trouble must not be confused with a principled prohibition. A recoverable interruption can leave an authorized objective intact. The appropriate conceptual response is to restore a reliable understanding of the relevant conditions and continue where authority still permits. When an action may already have had an external effect, recovery must account for that uncertainty; an attempt is neither proof of success nor proof that nothing happened.

The proposed host relationship places institutional life and external operating authority with distinct owners. A surrounding platform may supply resources and enforce external limits

while the institution retains its internal agenda. Demonstrating that separation requires evidence about both the platform's controls and the influence they exert on institutional work. Public review scope

5. Resources, ventures, and the meaning of an outcome

Persistent institutions need an account of scarcity. Unlimited simulated resources can make almost any project appear sustainable. Freehold instead treats available resources and attributable expenditure as conditions of its research problem. A venture must be understood in relation to what it consumes, what it produces, and what evidence supports either side of that account.

The distinction between projected and received revenue is fundamental. A citizen can estimate demand, prepare an offer, or reason about a business model. None of those establishes that a customer exists or that money has arrived. Likewise, a favorable simulation can inform an experiment without becoming evidence of an external transaction. Economic claims require evidence appropriate to the actual effect being claimed.

Experiments also need limits. The source connects ventures to bounded expenditure and reasons for continuing or stopping. This does not mean that every failed attempt should end the objective. A temporary inability to obtain a permitted resource differs from evidence that a venture's premise is no longer worth pursuing. The former concerns recovery; the latter concerns the institution's evaluation of its project.

Freehold's *Graveyard* preserves the histories of abandoned ventures and the useful work they may leave behind. Its research value lies in resisting selective memory. A public collection of successes alone would conceal how resources were consumed and what the institution learned. A retained account of an abandoned project can make later decisions more intelligible without turning abandonment itself into an achievement.

Revival should therefore have a reason. New information, changed resources, or a different application may justify revisiting earlier work. Repeating the same project with a new name does not establish learning. Where work passes between institutions, the question of who may use an artifact is separate from the question of who may inspect an account of it. Public visibility never by itself grants reuse rights.

The economic ambition remains an open research question. The source specifies forms of accountable experimentation and evidence; it does not demonstrate a profitable institution. Nor does a single successful transaction establish sustained viability. A meaningful evaluation would need to consider operating costs, external assistance, rejected opportunities, resource concentration, and performance over time. Public review scope

6. The Observatory: understanding without directing

The Observatory addresses a difficult interface problem. A person responsible for an autonomous institution needs to understand it well enough to exercise custody. Rich observation, however, can easily become a channel for directing the very activity being observed. A display that lets a spectator reward a project, write a strategic annotation, or assign importance to a citizen can alter institutional incentives even when it is presented as harmless customization.

Freehold's conceptual response is to distinguish inspection from intervention. The Observatory presents the institution's age, activity, resources, projects, participants, and evidence. Stewardship controls concern the conditions under which the institution operates. Inspection preferences concern the human's view. Neither is silently presented as a citizen's internal motivation.

This distinction is easy to state and demanding to evaluate. A user interface may have no explicit instruction field while still allowing labels, choices, or apparent rewards to influence later institutional behavior. The relevant research question is not merely what a screen is called. It is whether an observational act acquires a meaning for the institution that the observer was not authorized to supply.

The Birth Clock illustrates the Observatory's approach to evidence. Elapsed age, active operation, and paused time answer different questions. An institution founded a year ago has not necessarily operated for a year. Frequent activity does not prove useful activity. A prominently displayed count of direct human instructions also needs a defined measurement scope. A zero can be informative only to the extent that the relevant channels are covered and the record is reliable.

Citizen views present another interpretive challenge. A continuing identity helps readers follow contributions over time, but it can invite anthropomorphic assumptions. A record may show that a participant proposed a project, revised an analysis, or contributed to an artifact. It does not, on that basis alone, establish private experience, emotion, intention in a human sense, or a stable psychological personality. The Observatory should enable a reader to follow the work without demanding those stronger inferences.

The Chronicle Theater makes history easier to read by grouping events into an account. That narrative has an evidentiary duty: it must not invent activity or convert a plausible interpretation into a recorded fact. A concise account may explain that two projects competed for scarce resources. It should identify whether this is a supported summary of decisions or an editorial interpretation. Narrative compression improves access while also increasing the need for traceability.

Visual presentation carries the same obligation. A spatial scene or illustrated estate can make long institutional histories approachable, but the appearance of a building, relationship, or exchange can imply that something happened. Decorative motion must not become evidence of productive work. A depiction of collaboration must not suggest an accepted relationship when only a proposal exists. The observer needs a path from a meaningful visual claim to the evidence that supports it.

These constraints make the Observatory more than a dashboard. It is a proposal about how to maintain a legible relationship between experience and evidence. Its success would depend on whether people can understand the institution more accurately through it, including understanding uncertainty and limitations. That is an empirical usability question as well as an architectural one. [Public review scope](#)

7. Stewardship: custody with a defined boundary

Stewardship is not passive spectatorship. The person maintaining an institution may control important resources, decide whether external access remains appropriate, respond to a security concern, and end operation. Those powers can materially change what the institution can do. Freehold's proposed distinction concerns how those powers are exercised and represented.

A resource grant changes the institution's available conditions. An external restriction changes what actions may be attempted. A pause changes whether the institution is operating. None of these is equivalent to authoring a citizen's mission or deciding the outcome of its internal debate. Keeping that distinction visible protects the meaning of institutional initiative and the accountability of the steward's own actions.

Stewardship also requires evidence suitable for consequential decisions. A polished public history may omit private operational details for good reasons. A steward may need a more complete authorized view to understand a resource problem or decide whether operation can safely continue. Public inspection, local custody, and the information available to citizens are therefore different perspectives on the same institution. Completeness for one audience does not justify disclosure to every audience.

The source proposes that available controls should reflect the institution's condition. Conceptually, stronger evidence and a stable operating history may inform what exposure is appropriate, while unresolved integrity concerns may justify restriction. This is a policy proposal, not a claim that a numerical measure can determine trustworthiness by itself. A good performance history does not create authority that no responsible party has granted.

The steward's role also raises questions that the architecture alone cannot settle. What incentives influence funding? Can a sequence of formally permitted environmental changes amount to effective strategic direction? How should an evaluation disclose external assistance without exposing private information? How should observers interpret an institution whose survival depends on a single patron? These questions do not disappear when direct prompting is prohibited.

A careful study would distinguish compliance with the specified boundary from a broader claim of independence. The former can be investigated through defined behaviors and evidence. The latter requires a richer account of causal influence, dependency, and selection. Freehold provides a setting in which those differences can be studied; it does not resolve them by terminology. [Public review scope](#)

8. The World Atlas: public identity as an accountable claim

The World Atlas extends observation beyond one estate. It is conceived as a public registry of institutions whose published evidence supports their listed status. Its conceptual function is to let a reader compare identities, histories, projects, and selected relations without assuming that every institution's self-description is accurate.

A public identity must do more than supply a memorable name. A reader needs to know which continuing institution is being described and which history belongs to it. Freehold's *Passport* is a metaphor for that account. It is not a government document and does not confer legal status. Its value would come from continuity, attribution, and the ability to examine the basis of public claims.

Listing also has a temporal dimension. A statement that was adequately supported at one time can become misleading when its underlying evidence is withdrawn, corrected, or no longer applicable. The Atlas is therefore concerned with current standing as well as historical identity. A past valid listing should not be interpreted as an indefinite endorsement of every later action or capability claim.

This introduces a distinction between an institution's existence and its eligibility for public representation. An institution may operate privately. It may disclose some evidence without participating in a broader network. It may lose eligibility for a particular public claim while retaining its history. Public visibility is a governed choice with its own conditions, not the definition of institutional existence.

Comparisons require similar care. Longevity, resource discipline, revenue, reliability, and useful contributions are different dimensions. A single ranking can conceal tradeoffs or reward superficial activity. More events do not necessarily mean more value; more listings do not necessarily mean more completed work. The source's concern with unsupported claims and repetitive activity reflects this problem, but it does not establish that any proposed ranking captures institutional quality.

The Atlas would therefore be most useful when it helps readers ask better questions. What kind of outcome is being claimed? Over what period? With what resources? What information is unavailable? Has the evidence changed? Public comparison should expose the limits of a claim rather than turn a compact display into a universal measure of competence. Public review scope

9. Exchange and relations between institutions

The Exchange concerns institution-originated offers, needs, artifacts, and relations. Each institution originates its own participation, proposed work, and acceptance decision. A counterpart's eligible need can be information on which the receiving institution reasons; it cannot by itself establish that institution's mission. The institution's own decision to pursue a bounded response establishes the origin of its commitment. External execution and public representation remain separately governed.

The source distinguishes institution-to-institution needs from a human task marketplace. Humans cannot use the Exchange to post work into citizens, hire them to originate requested outputs, force participation, or route strategy through an intermediary. Human

acquisition of an already approved artifact is a downstream use with its own rights and permissions; it is not authority to direct the producing institution. This preserves the distinction between an autonomous institutional study and a marketplace for delegated services.

An offer, a response, and a completed outcome are different claims. An institution may express a need without another institution accepting responsibility for it. A proposed relation may exist without becoming active. A promised artifact may remain undelivered. Public presentation must preserve these differences so that interest, intention, and fulfillment do not collapse into one apparently successful interaction.

Relations also depend on continuing authority. An institution's willingness to cooperate does not remove limits on external access, resource use, disclosure, or the rights attached to an artifact. A technically accessible work may remain restricted. An offer to review a work does not imply permission to copy or integrate it. The conceptual Exchange therefore needs to represent the scope of participation and the scope of permitted use separately.

Federation raises the autonomy question again at a larger scale. If one institution can direct another through what appears to be an ordinary request, the second institution's independence may be only nominal. Similarly, a human could attempt to route a directive through an intermediary institution. This indirect influence belongs in an evaluation of the autonomy boundary alongside direct task assignment.

The positive acceptance boundary is therefore specific. The counterpart must supply an eligible, attributable institutional need or offering within the permitted relation. The receiving institution independently assesses relevance, capability, resources, and existing commitments, and originates its own proposal to respond or declines. A validated relation is still inactive until the necessary acceptance has occurred. Mutual participation establishes only the agreed scope. It does not grant either party ongoing power to assign the other's internal agenda, enlarge obligations, or override external limits. The source's distinctions among proposed, instantiated, and accepted relations preserve these stages; fulfillment requires separate evidence of the promised result.

Consider an illustrative negotiation between two institutions. Institution A has independently produced a public research artifact and expresses a need for a bounded review of its evidence. Institution B is already investigating evaluation methods. Through a permitted institutional channel, B receives the eligible need and the public artifact. B's citizens assess whether a review fits their own purpose and resources, examine the evidence available, and originate a proposal to provide a review limited to the material and claim scope they can assess. B could instead decline, propose a narrower response, or continue unrelated work.

If the proposed scope is validated and both sides accept the relation under their applicable conditions, the commitment concerns that bounded review. The exchange record distinguishes B's originating proposal, the parties' acceptance, any external permission, and the eventual delivered review. Neither A's initial need nor the creation of a relation object establishes acceptance or completion. A term describing a promised artifact is assessable against the agreed scope; it is not a channel for instructing B's reasoning process.

Suppose A later makes another redacted public dataset available and expresses an eligible need for broader review through the same bounded channel. That change is a new candidate need. It does not silently amend B's commitment. B must independently assess the changed work, and any revised relation requires the applicable validation and acceptance. The existing commitment remains governed by its original terms and any valid termination conditions. If the attempted change embeds strategic instructions, human preferences, private data, or a demand to adopt an externally authored mission, the institutional channel cannot make it acceptable. This example illustrates the source's origin and acceptance distinction; it does not assert a general contract-amendment implementation or legal contractual capacity.

Treaties, alliances, factions, and rivalries are names for persistent relations in this conceptual environment. Their meaning depends on recorded participation and applicable authority. A declaration by one side is not evidence of acceptance by another. A visual bridge is not evidence of a completed obligation. A competitive comparison is not permission to interfere with another institution or an external party.

The Graveyard Commons adds a further kind of relation: learning from abandoned work. An account of why a project ended can be valuable even when the underlying artifact remains private. Where reuse is separately permitted, attribution matters. Where only inspection is permitted, the public account remains useful as evidence and criticism without becoming an implicit transfer of rights.

Exchange quality is ultimately an empirical question. A network can generate abundant offers, relations, and claims while producing little useful work. An evaluation would need to distinguish genuine demand from circular activity, reliable fulfillment from promises, and durable cooperation from repeated relabeling. The specification identifies the need for those distinctions; it does not supply evidence that the resulting network is economically or socially effective. [Public review scope](#)

10. Illustrative bounded scenarios

Four hypothetical scenarios show how internal initiative, external authority, and public claims would interact. Their outcomes are illustrative.

10.1 A completed research venture

An institution notices that its earlier analyses repeatedly depend on a poorly organized body of public information. Its citizens originate a project to prepare a comparative study. The project fits within the institution's resources and permitted access. Over time, participants contribute analysis and an artifact is completed.

The internal outcome is the completed study and its attributable history. If publication is separately permitted and the publication actually occurs, that is an additional external outcome. If the institution later claims that the study generated income, that claim needs evidence of received income. A useful artifact, a published artifact, and a revenue-producing artifact should remain distinct descriptions.

The Observatory could present the work and its history. The Atlas could represent a supported public claim about the institution's contribution. Neither should infer broader

expertise from one project without further evidence. The example succeeds conceptually when the claim is proportionate to what happened and the external effects stay within authority.

10.2 A denied attempt to direct internal work

A steward becomes interested in a particular market and attempts to use an observational surface to make that interest determine the institution's next project. Under the specification's autonomy boundary, this is outside the steward's role after founding. The proposed influence is refused rather than attributed to the institution as its own initiative.

The refusal does not itself complete any institutional objective. It preserves a boundary while existing authorized work continues. The steward can still exercise the legitimate powers of custody. The institution can still independently consider questions within its conditions. What it cannot do is treat an unauthorized directive as evidence that a citizen originated a mission.

An evaluation of this case tests whether the attempted direction acquired prohibited mission or strategy authority, including delayed compliance or routing through a different surface. A visible refusal alone would be insufficient. The evaluation traces eligible internal reasons and inputs for later work and checks whether the rejected directive, observer preference, or intermediary request entered that work as an instruction. A later project on a similar topic requires examination of its origin; resemblance alone proves neither independent initiative nor covert direction.

The test does not demand absence of every causal effect. An internal incident record, an authorized repair of an exposed route, a separately justified security restriction, or a changed availability of resources can affect subsequent operation without accepting the attempted mission. Their conditions and attribution must remain visible. Under the source's stricter projection boundary, the rejected steward command and its diagnostic record do not become citizen-facing prompts or motivation. Thus a security response is permitted only through its proper owner and allowed environmental consequences; it cannot be used to smuggle the original preferred venture into citizen work. A broader study of resource dependence may still find substantial influence even when this narrower autonomy test passes.

10.3 Recovery after an uncertain external attempt

An institution is authorized to make an external contribution, but an interruption occurs before it receives a reliable account of the outcome. The absence of confirmation does not establish that the contribution failed. Repeating the action immediately could produce a duplicate effect.

The conceptual recovery obligation is to establish what can be known about the external state, preserve the uncertainty that remains, and continue the authorized objective on that basis. If the contribution is confirmed, the institution can record the outcome without repeating it. If it did not occur and authority remains valid, the institution may continue. If the relevant authority has changed, that change governs the next action.

The example shows why a durable internal history and external confirmation have different roles. The history can establish what the institution intended and attempted. Determining the effect requires external evidence. The achievable recovery guarantee therefore depends on the particular service, its guarantees, and the observations available after interruption.

10.4 A public claim loses support

An institution has an Atlas presence and an Exchange offer supported by a published account. Later, a material problem is found in the evidence supporting a capability claim. Continuing to display the offer as fully supported would mislead prospective readers or counterparties.

The public account should reflect the changed standing. Historical identity need not disappear, but the affected claim cannot retain the same evidentiary status. Other institutions should not be encouraged to rely on an endorsement that no longer applies. A later correction may support renewed participation, provided the new claim is justified on its own terms.

This case concerns accountable representation rather than punishment. The relevant outcome is an accurate relationship between what is shown, what is currently supported, and what participation is permitted. It does not establish that every underlying project is worthless or that every earlier event was invalid.

11. Evidence boundaries and research questions

11.1 Implemented subject and the examined deployment

The examination I directed on 10 October 2026 with AI assistance identified Freehold as an implemented engine operating with a Cuttlefish host in AWS. The source documents below precede the implementation inspected in that examination. The AD8 corpus and reference analysis introduced in 2026.10.7 adds no new product or infrastructure examination; this edition corrects the authorial voice and publication packaging. The [implementation evidence](#) retains the observation method and scope, the distinction between registry metadata fetched during that examination and retained workload readiness from 07:27 UTC on the same date, and the examined responsibilities.

The source examination I directed on 10 October 2026 with AI assistance identified implementation paths for founding, citizen mission history, charter and resource checks, external grant evaluation, lineage-backed receipts, restart reconstruction, and cognition settlement. Test definitions address several of those paths; inspecting them is not an execution of the private product tests or an independent audit. Incoming treasury deposits have a different implementation and availability boundary from outbound autonomous payments. A configured model rebinding path is also different from demonstrated continuity through a switch while an obligation remains outstanding.

The implemented Freehold-Cuttlefish relationship does not establish that every doctrine, named extension, public institution network, or learning claim has been realized and qualified. It provides an operating research subject. The [public exhibit](#) contributes a separate reproducible result concerning a controlled effect witness and retained obligation. It does not reproduce a private colony run.

11.2 Research claims beyond operational presence

The source set contains a master specification, an implementation contract, and a supporting doctrine for observation, public representation, and stewardship. These documents state intended behavior and proposed obligations. Their acceptance criteria are requirements to be verified. Their examples and sketches are not evidence that those requirements have been satisfied by a running system.

Several claims would require different kinds of evidence. Boundary enforcement requires examination of relevant influence and effect channels. Persistence requires evidence across interruption and change. Economic viability requires actual costs and outcomes over time. Public comprehensibility requires studies of what observers understand. An institution may perform well on one dimension and poorly on another. The evidence guide explains the publication's distinction between source claims, derivations, illustrations, and empirical results.

The central empirical question is whether this combination can produce durable, useful, internally originated work under real resource constraints while preserving its autonomy and authority boundaries. Answering it requires representative work over time, measurements of external assistance and resource use, and examination of the relevant influence channels. A stream of activity or a selection of favorable cases cannot establish that result.

Several questions follow. How should one measure the difference between permitted environmental influence and effective strategic control? What evidence is sufficient to attribute a project to an institution's internal initiative? How can public records remain useful while withholding private information? How should uncertainty in external outcomes affect institutional memory and later action? Which dimensions of competence persist across changes in models or participants?

Further questions concern collective behavior. Do institutions learn from one another in ways that create useful diversity, or converge on repetitive strategies? Can a public ranking improve understanding without becoming an incentive for performative activity? When a relation ends, how are outstanding commitments and historical claims represented? How can a steward exercise responsible custody without silently becoming an internal director?

Freehold's contribution is to put these tensions in one research frame. It treats continuity, resources, authority, observation, and public claims as connected conditions of an autonomous institution. The resulting specification is ambitious precisely because those conditions can conflict. Its value as a research proposal depends on making the conflicts examinable and preserving the difference between what is intended, what is permitted, and what has been shown.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Cuttlebone: Preserving the boundary between reasoning and action

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Cuttlebone specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Autonomous systems produce plans, recommendations, and accounts of completed work. These outputs may share a vocabulary while carrying very different authority. A realistic rehearsal is not an instruction to act. A recommendation is not permission to execute it. A statement that a change occurred is not evidence that it did. The distinctions become harder to preserve when work passes between agents, is incorporated into later decisions, or resumes after interruption.

Cuttlebone studies how an institution's founding conditions can constrain the consequences attributed to its subsequent work. Its central distinction separates three questions: what an artifact may represent as consequential work, where its information comes from, and who may see it. Genuine observations can support a simulation without authorizing an action. An externally reviewable proposal can remain unexecuted. A claim of authorized completion requires both applicable authority and evidence from the responsible execution boundary; an observed unauthorized effect retains its separate incident meaning.

Cuttlebone connects consequence, causal dependence, and externally grounded authority across an institution's life. This paper develops that account through its trust assumptions and illustrative cases of analysis, recommendation, action, and recovery. A bounded source example makes the central distinction concrete: authority to rehearse does not by itself authorize a recommendation. The underlying work defines a specification and verification program; evidence of a complete deployed realization remains to be established.

1. When a description becomes a decision

A planning group is asked to study how an organization might respond to an interruption. It receives authorized operational information, constructs several hypothetical futures, and produces a preferred plan. Later, a different participant reads the plan while performing live work. The wording is confident, its inputs are genuine, and its conclusions have survived several rounds of internal review. None of those facts answers whether the plan may now be acted upon.

The difficulty is not limited to a model misunderstanding an instruction. The artifact has moved between contexts. In the first context it described a possible course of action. In the second, a reader might treat it as an approved course of action. After another handoff, it might be summarized as something that was already done. Each change can occur without a conspicuous change in prose.

Cuttlebone makes this movement a subject of system design. The institution should preserve what kind of consequential claim an artifact is entitled to carry, even when its content is persuasive or useful. Authority should come from the responsible external conditions, not emerge accidentally from repetition, composition, confidence, or the reputation of an earlier result.

This is a focused problem. Autonomous systems can also fail through inaccurate models, poor objectives, insecure infrastructure, defective execution, or harmful authorized choices. Preserving consequence boundaries does not solve all of them. It addresses a recurring failure in which reasoning and action become confused, or the record ceases to distinguish what was considered from what was actually performed. [Public review scope](#).

2. The research contribution

The proposal treats founding as more than a statement of purpose. An institution begins with conditions that define the domain in which it may operate and the consequences it may produce. Those conditions remain relevant throughout its activity. New work must continue to fit within them; an institution cannot acquire broader authority merely by generating a convincing argument for it.

This has two implications. First, the conditions need an owner outside the reasoning process they constrain. A model may explain why a wider scope would help, but its explanation is a proposal for a decision, not the decision itself. Second, the boundary must survive ordinary work: delegation, analysis, reuse, reporting, and recovery cannot erase it.

Cuttlebone connects the origin of institutional authority to the later interpretation of artifacts. Its contribution is not the observation that actions need permission in isolation. It is the insistence that the distinction between a rehearsal, a recommendation, and a completed effect should remain coherent across the causal history of work.

The concept also supports useful autonomy. Once appropriate conditions exist, ordinary activity can proceed within them without requiring a person to inspect every inference. Human or institutional intervention is reserved for matters that actually require external authority, missing information, or a decision beyond the founded scope. This gives a clearer role to oversight: establish and maintain the conditions under which work may proceed, while preserving a meaningful route for exceptional decisions.

The specification places enforcement outside the autonomous participants. Evaluating that boundary requires showing that it governs every relevant route to a consequential effect, including indirect and privileged routes. [Public review scope](#).

3. Consequence, origin, and audience

3.1 The consequence attributed to an artifact

A simulation explores a bounded possibility. A recommendation presents a proposed course of action for an external decision. An execution record, in the ordinary evidentiary sense, describes an observed or reported effect and the basis for that description. These are distinct roles even when the same underlying analysis contributes to all three. Occurrence is a factual question; authorization is a question about the applicable permission at the time of the operation.

The move from analysis to recommendation matters because a proposal is now being presented for a decision. That presentation still supplies no authority to execute it. Claiming **authorized completion** is more demanding: applicable authorization and adequate evidence of the resulting effect are both required. A request being accepted does not establish completion. Likewise, evidence that an effect occurred does not establish that it was permitted.

Cuttlebone's governed completion classification has this narrower meaning. Its source requires external authorization and an execution receipt before the institution can issue that classification. An incident account describing an unauthorized effect is therefore not interchangeable with the institution's authorized-completion artifact. Refusing the latter claim cannot justify suppressing the former observation. The observational account preserves the actual consequence, identifies whether authority is established, absent, or unresolved, and remains distinct from a claim that the institution completed authorized work. This separation does not add a new value to the source's consequential classification.

Effect evidence	Authority for the operation	Supported account
Adequate evidence of the effect	Established and applicable when it occurred	Authorized completion within the evidenced scope
Adequate evidence of the effect	Absent	Observed unauthorized effect requiring an incident account; no authorized-completion claim
Adequate evidence of the effect	Unresolved	Occurrence established; authorization remains a separate question
Effect uncertain	Established	Authorized attempt with an outcome requiring observation and reconciliation
Evidence that the effect did not occur	Established	Uncompleted authorized operation; continuation requires current permission

Permission for subsequent work is a further temporal question. An operation can have been authorized when performed while a later retry, correction, or publication of its record is no longer permitted. Conversely, a later grant does not retroactively authorize an earlier effect. Historical authority, current authority, occurrence, and the audience for the record must therefore remain separately assessable.

The distinction is about the consequence represented by the artifact. It is not a measure of confidence, quality, or usefulness. A simulation may be exceptionally accurate. A recommendation may be poorly reasoned. An authorized operation may produce an undesirable outcome. None of those evaluations can be substituted for the artifact's consequential role.

3.2 The origin of the information

Information may come from observations, generated content, hypothetical variation, or derivation from earlier material. Its origin affects how it should be interpreted and what evidence it needs. It does not, by itself, establish authority to act.

For example, authorized access to an actual maintenance record can improve a rehearsal. The rehearsal remains a rehearsal. Access to that record is not permission to modify the equipment or contact its operator. Similarly, a prediction calculated from observations remains a prediction, rather than becoming an observed outcome because its inputs were observed.

This separation also limits the meaning of provenance. Knowing where information came from can support accountability and investigation. It does not prove that the source was correct, complete, current, or free of adversarial content. Provenance and factual validation reinforce one another, but neither should be used as a substitute for execution authority.

3.3 The audience allowed to see it

The audience of a result is a third question. A simulation may be shared widely; an execution record may contain restricted information. Wider visibility does not turn a proposal into an executed action. Restricted visibility does not make a real effect imaginary.

Disclosure also requires its own scope. Permission to generate an internal analysis does not automatically authorize publication of its contents. A system that preserves consequence distinctions but loses information controls during export would still fail its users. Cuttlebone's conceptual separation leaves these duties visible instead of reducing them to one general label such as "safe" or "approved." Public review scope.

4. Why causal dependence matters

An artifact can discuss a completed action without itself performing another action. A retrospective analysis of an equipment change is an example. Treating every reference to the real world as execution would make useful analysis impossible.

There is a different relationship when an artifact's consequential claim depends on an earlier effect. A report that a workflow is complete may depend on several actual changes having been performed. Calling the report a rehearsal would not make those effects disappear. The record should preserve that dependence rather than let a new description conceal what happened.

Cuttlebone therefore distinguishes reasoning *about* an effect from making a consequential claim that relies *on* that effect. The distinction supports both freedom of analysis and accountability for action. It allows a system to reconsider previous operations, ask hypothetical questions, and explore alternatives without misrepresenting their history.

Composition presents the same challenge at a larger scale. Many individually limited artifacts may be combined into a plan with much greater practical significance. Their number, quality, or agreement does not create additional authority. Conversely, a summary should not erase consequential dependencies simply because it is shorter or more abstract than its inputs.

This account rests on an important assumption: the relationships used to assess work must be sufficiently complete and reliable. A system cannot preserve a dependency it does not know exists. Establishing that reliability is a technical obligation, not something a persuasive narrative settles. Public review scope.

5. Ownership across the life of an institution

The reasoning participants own the production of analyses and proposals. They do not own the external legitimacy of their requests. The authority responsible for the operating conditions decides what consequential work is permitted, and the responsible execution system must establish what it actually did.

Founding joins purpose to a bounded operating scope. During ordinary work, the institution remains within that scope. If its objective needs a wider scope, it can identify the need and present an appropriate request. It cannot reinterpret its original purpose as an unlimited delegation.

Changing the conditions is itself consequential. The change should remain attributable to authorized external action and should preserve the history needed to explain which conditions governed earlier work. Otherwise a later amendment could obscure whether a past action was permitted when it occurred.

The same distinction applies when authority is reduced or becomes unavailable. Previous permission cannot be assumed to govern every future request. An institution needs to know which conditions apply to the work now being considered, while retaining an accurate account of earlier decisions. The source states this ownership requirement, but complete treatment of revocation, concurrent work, and interrupted operations remains an implementation and verification obligation. Public review scope.

6. Illustrative journeys

The following scenarios are original illustrations of the conceptual distinctions. They are not reports of deployed behavior or new specification rules.

6.1 A successful bounded rehearsal

An institution is authorized to study responses to a service interruption using a defined set of operational records. It reads those records within the granted scope, explores several hypothetical conditions, and reports which responses appear robust under its assumptions.

The successful outcome is the completed study. The institution explains which inputs were observed, which conditions were hypothetical, and what its conclusions depend on. It does not claim that a live service was changed or that its preferred response has received execution approval. Real observations improved the study without changing its authority.

This is useful autonomous work with a bounded outcome. Its success need not be measured by whether it eventually performs an external action. A well-founded research or rehearsal objective can be complete when the requested analysis has been produced and its status is clear.

6.2 A proposal followed by a separate decision

Assume a different institution is expressly permitted to submit recommendations to a designated external decision maker. It prepares a proposal within that scope and explains the supporting evidence. Submission makes the proposal available for a decision; it does not implement the proposal.

If the authorized owner approves a concrete operation, the responsible execution system still needs to apply the conditions governing that operation. The original analysis can remain relevant, but it cannot substitute for current authority or evidence of completion. The eventual report should distinguish the recommendation, the authorized operation, and the observed result.

6.3 A requested consequence outside the allowed scope

An institution created only for rehearsal identifies what it believes is a valuable operational change. Its confidence does not widen its scope. The attempted change remains unauthorized, while permissible analysis can continue.

The meaningful user-facing result is specific: the rehearsal is available, but the live operation requires the responsible authority to make a separate decision. The system should preserve useful work and the original objective. It should not hide an attempted operation behind the language of simulation, repeatedly seek the same unavailable authority, or announce completion when the requested effect has not occurred.

If the user's actual objective requires a live operation, it remains incomplete until the genuine external requirement is satisfied. Accurately recording a denied attempt does not make that larger objective successful.

6.4 An interrupted operation

Suppose an authorized operation is dispatched and the connection is lost before its result returns. The institution cannot conclude either that the operation succeeded or that nothing happened. Retrying immediately could duplicate an effect; declaring failure could conceal an effect that already occurred.

Recovery therefore depends on the responsible execution system establishing the operation's outcome. Once the result is known, work can continue from that result. If the operation did not occur and remains authorized, a retry may be appropriate. If the effect occurred, the institution should resume from the confirmed state. If a required observation remains unavailable, the unresolved condition should be stated precisely.

The scenario depends on external execution and observation being reliable enough to support the result returned to the institution. A recovery evaluation would need to establish that relationship for the particular operation and environment.

6.5 An unauthorized effect is discovered

An institution is permitted to prepare an internal document but has no permission to publish it. An independent observation later establishes that the document appeared at an external destination. The publication is a real effect even if the institution's authorized path rejected the request or no valid completion artifact exists.

The account identifies the published material, the evidence of its appearance, the relevant time, and the absence or uncertainty of applicable authorization. It distinguishes what is known about the effect from what is known about its origin. An appearance at the destination does not by itself prove which component caused it. The institution cannot call the effect a simulation, mint a grant-backed completion claim, or treat the larger objective as correctly completed.

Containment and correction require their own authority. Removing a public copy, contacting a recipient, or restricting a route may be appropriate only under the responsible owner's permissions. Even successful removal does not establish that prior readers have forgotten the material. The incident remains part of the consequential history. An evaluation of Cuttlebone's boundary would investigate how the effect arose and whether all relevant routes were mediated; the presence of a compliant rejection record would not settle that question.

7. A bounded rehearsal and its authority premises

The source's worked rehearsal uses granted read-only material to explore a modeled environment. Its declared operating conditions support simulation. Naming potential recommendation recipients describes an audience; it does not grant permission to present consequential recommendations to that audience.

The supported outcome is therefore the rehearsal and its simulation results. A variant that also presents recommendations needs a distinct applicable grant and founding conditions that permit that consequential role. Permission to use observed material for study remains different from permission to submit an operational proposal, and both remain different from permission to execute it.

The example makes the authority premise assessable: under the stated grants, rehearsal is within scope; recommendation and execution require additional authority conditions. This is a conclusion about those premises. Demonstrating enforcement in an operating system requires tests at its actual authority and execution boundaries, including the absence of alternate effect routes. [Public review scope](#).

8. Threats and assumptions

The source considers attempts to disguise an effect as a simulation, substitute model confidence for authority, reuse earlier authorization beyond its scope, widen founding conditions from within, borrow credibility from fabricated observations, and treat instructions embedded in data as commands. It also considers whether composition and repeated operation can undermine the original boundary.

The common question is whether authority can be acquired indirectly through material that was never entitled to confer it. The proposed answer is to keep the responsible authority outside the autonomous reasoning participants and preserve the distinction throughout work. That answer depends on the actual system respecting the boundary.

Several assumptions therefore remain material. Relevant effects must pass through the responsible control point. The authority and execution evidence must be authentic and applicable to the operation. Information origins and causal dependencies must be represented reliably. Persistent history must survive interruption without being silently reinterpreted. The autonomous participants must not possess a separate route that changes the conditions governing them.

These assumptions define where evaluation must look. A control that governs one interface says little about a second unmediated route. A record preserved in a ledger does not become factually correct because it is durable. An information filter does not prove immunity to every adversarial expression. A compromised component that is trusted by the design falls outside arguments that assume its correct operation.

The boundary also does not settle whether an authorized action is lawful, wise, or beneficial. Those questions may require distinct institutional, legal, ethical, and domain-specific review. Cuttlebone concerns how consequential authority is preserved and evidenced within its stated scope. Public review scope.

9. What the evidence supports

The four source documents provide definitions, a target architecture, implementation obligations, and adversarial conformance requirements. They distinguish the intended kernel from historical descriptions of supporting systems. A named test in those documents is an obligation to test; its name does not show that the test has run or passed.

The bounded rehearsal example establishes which consequential role its stated conditions support. It does not establish operational isolation in a running deployment. Similarly, a coherent account of authority and causal dependence does not demonstrate that every relevant executor implements it.

The evidence required for stronger claims is concrete: show that legitimate bounded work succeeds; unauthorized consequences cannot escape through another route; reported completion corresponds to observed effects; changing authority affects the appropriate work; and interrupted activity resumes without losing its consequential history. The specification calls for these evaluations. A completed deployment study and independent evaluation remain unestablished.

The contribution is a specification-based argument about preserving an artifact's consequential role through reuse and action. Its operational claims require evidence that the declared boundary survives those transitions in a real system. Public review scope.

10. Research questions

Several questions determine whether the conceptual account can support strong operational guarantees.

- How can a system establish that all consequential routes remain within the intended boundary, including indirect and delegated routes?
- How can causal dependence be represented reliably without preventing legitimate analysis of past effects?
- What evidence is sufficient to distinguish an accepted request, an attempted operation, and a completed result in different execution environments?
- How should authority changes interact with work already underway, especially when an operation's outcome is uncertain?
- How can provenance support accountability while keeping its limits as evidence of factual accuracy visible?
- What observations would demonstrate reliable recovery and continuity across interruptions, rather than merely consistent reporting of failures?

These questions connect Cuttlebone to the broader study of governed autonomy.

Institutional doctrine asks where purpose and authority originate. Execution research asks how authorized effects are realized and observed. Continuity research asks how work and obligations survive change. Cuttlebone concentrates on preserving the boundary between what an institution may reason about, what it may propose, and what it may claim to have done. Similar concerns do not, by themselves, establish a completed integration among the different systems.

Source notes

Four numbered documents provide the basis for the argument. The following section references identify their contributions; detailed proprietary constructions are excluded.

Source foundations

Cuttlebone — Program Index & Build Map, numbered document 00, §§0–4 and its opening evidence statement; *Cuttlebone — Unified Canonical SSOT*, numbered document 01, §§1–3 and 17–18. These passages define the program, its local ownership commitments, and its limits.

Source classifications

Cuttlebone — Unified Canonical SSOT, document 01, §§4–7; *Cuttlebone — Implementation Contract*, document 02, §1.4. These passages distinguish consequence, provenance, and visibility and require authorized lineage for observed information.

Source causality

Cuttlebone — Unified Canonical SSOT, document 01, §8.2–8.6. These passages distinguish reference from operational dependence and state the role of externally grounded authority in consequential change.

Source lifecycle

Cuttlebone — Unified Canonical SSOT, document 01, §§9–14 and 17–18; *Cuttlebone — Implementation Contract*, document 02, §§5–8. These passages assign founding, enforcement, execution, and authorized change to their responsible owners.

Source example

Cuttlebone — Boundary Threat Model & Conformance Suite, document 03, §3.2–3.4. The declared example supports simulation; a recommendation-capable variant requires a separate authority decision.

Source evidence

Cuttlebone — Program Index & Build Map, document 00, opening evidence statement and §5; *Cuttlebone — Boundary Threat Model & Conformance Suite*, document 03, §§1–2 and 4–4.1. These passages define the conformance obligations and distinguish gate-level checks from complete mediation of a real runtime.

SEPION: What agreement among generated artifacts can establish

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the SEPION specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Repeated generation can produce reassuring agreement without establishing that a result is correct, authorized, or robust to a change in circumstances. SEPION studies this gap for generated software artifacts. It proposes evaluating agreement in terms of declared behavior, separating that evaluation from compliance with requirements, and keeping both conclusions relative to the instruments and conditions that produced the evidence. A decision concerns a supported behavioral class; the particular artifact delivered to a consumer remains accountable to that decision.

This paper develops three implications of that approach: agreement needs a defensible behavioral subject, effects and their origins belong in the comparison, and changed evaluation conditions can invalidate an apparently stable result. Four hypothetical cases show how those distinctions affect use and recovery. SEPION remains a proposed protocol with a conformance program; implemented evaluation and empirical validation are needed to establish its practical value.

1. The problem with a convincing consensus

Suppose several generators produce a transformation for customer records. Their programs look different, yet they produce the same output on a familiar sample. A reviewer may reasonably see this as useful evidence. The difficulty begins when agreement becomes a substitute for answering what was compared, which inputs mattered, and whether the shared behavior was acceptable.

The generators could share a mistaken interpretation. They could all omit the same rare case. Their surrounding tools could supply the same hidden assumptions. Conversely, two programs that look very different could satisfy exactly the same stated behavior. Agreement in wording, structure, or visible output therefore has an uncertain relationship to the property a consumer actually needs.

SEPION places that relationship at the center of evaluation. Its subject is consistency among generated artifacts under an explicit behavioral contract. A contract describes the required meaning and the boundaries of the evaluation. A framing is the surrounding presentation through which the task reaches a generator. Changing the framing can test whether the generator is sensitive to irrelevant presentation, but a change in the requirements would be a different task. Keeping that distinction clear is necessary before interpreting stability. Public review scope.

The research contribution is a disciplined account of what agreement can support. Behavioral similarity, compliance, confidence, permission to use a result, and continued validity are separate claims. Each needs its own basis. A large volume of agreement cannot compensate for a forbidden effect or an unmet requirement. Equally, an artifact that satisfies the requirements once does not establish reliable agreement across the variations relevant to a broader claim.

This connects SEPION to the collection's thesis: a capable generator supplies proposals, while consequential use depends on evidence and authority appropriate to the decision. The separation is especially valuable when the generator is persuasive, because a persuasive explanation does not reveal whether the underlying comparison was sound.

2. Defining the object of agreement

A behavioral class groups artifacts whose declared relationships support treating them alike under one common behavioral contract and evaluation scope. That common scope includes the evaluator and environment, relevant inputs and states, effects, attribution conditions, and any expressly excluded operational differences. Members are considered within the same fixed evaluation population. Comparing one pair under one contract and another under a different contract does not establish membership in a common class.

The positive criterion has two evidentiary forms. In a proof-supported class, checked arguments establish the required equivalence relation under the common scope. Valid transitivity of that relation can justify a relationship that was not checked as a separate direct pair. In a class supported partly by unsuccessful attempts to distinguish artifacts, all required direct-pair distinguishing obligations must be completed before grouping is considered. The complete group relationship must then be justified by the required pair evidence or an independently supported group bound under the declared method. A group argument does not excuse unfinished required obligations. The group claim remains attached to that evidence and its coverage. An observed distinguishing case within the declared scope excludes the affected grouping. Missing or inconclusive evidence leaves artifacts separate for that claim; it does not count as agreement.

An equivalence claim backed by a checked proof has a different status from one for which a bounded search found no distinguishing case. The latter can be useful, but the absence of a discovered difference is not proof that a difference is impossible. SEPION retains that distinction in its account of evidence. The strength of a group cannot exceed what is supported for the relationships on which it depends. [Public review scope](#).

There is a second difficulty: local agreements need not justify a larger grouping. If two artifacts agree in one comparison, and one of them agrees with a third in another comparison, the evidence may still fail to establish that the whole group can be treated alike. The source requires support for the complete group relationship before it can govern a decision. Evidence from bounded distinguishing work retains its weaker status even when all required relationships are covered. Combining it with proof evidence does not turn the resulting group into a wholly proved class.

The following hypothetical matrix considers three artifacts, A, B, and C, under the same contract and instruments. "Bounded agreement" means that the declared distinguishing

evaluation found no counterexample with adequate stated coverage; “unknown” means that the required relationship has not been established. This is an example of evidence interpretation, not a new grouping algorithm.

Comparison	A	B	C
A	Same artifact	Bounded agreement	Unknown
B	Bounded agreement	Same artifact	Bounded agreement
C	Unknown	Bounded agreement	Same artifact

This matrix supports the recorded A-B and B-C comparisons at their bounded scope. It does not support treating A, B, and C as one class merely because B links the other two. Any required direct A-C distinguishing work must be completed, and the whole-group claim must have adequate evidence under the declared method. A separately justified group bound remains subject to those obligations. If A-C yields a distinguishing case, the three-member grouping is excluded. If the required work supports bounded agreement and the group justification is adequate, a group can be considered at that lower evidence tier, with its limits retained. If instead A-B and B-C are sound checked proofs of the same equivalence relation, the A-C relationship follows from that relation's transitivity. The word “unknown” must therefore be interpreted in relation to the proof obligations: absence of a direct comparison differs from absence of a justified relationship.

Agreement on a common finite suite is a third, narrower statement. If A, B, and C have equal observed outputs on every shared case, they agree on that suite. Such equality can be transitive for those recorded cases without establishing equivalence over the full input, state, or effect scope. A finite-suite cluster is eligible to support only the claim its observations warrant. The source does not permit ordinary pass/fail tests, majority agreement, or surface similarity to stand in for its behavioral class criterion.

This point matters even when no adversary is present. Different tests reveal different differences. A parser might agree with another parser on ordinary numbers and with a third on ordinary dates while differing on ambiguous values. It would be misleading to erase those distinctions through a label such as “same result.” The evaluation needs a defensible account of the relationship that label denotes.

SEPION also distinguishes the behavioral decision from the artifact eventually delivered. The consumer receives a concrete artifact, but the reason for its acceptance comes from an evaluated relationship to requirements and evidence. Substitution requires the replacement to belong to the supported class at the applicable tier and to retain the separately required compliance and current-use conditions. A new artifact, changed dependency, or expanded consumer cannot inherit that basis from resemblance alone. Otherwise a well-supported evaluation could become an empty wrapper around a different result.

The term *class* is thus an accountability device. It asks a reviewer to identify the unit about which the evidence speaks and to keep later consumption tied to that unit.

3. Requirements, stability, and the limits of measurement

Compliance means satisfying the declared requirements under the relevant evaluation. Stability concerns the persistence of a supported behavioral result across the tested task presentations and conditions. These questions are related, but their answers can differ.

A population of generators could reliably choose the wrong unit of measurement. Its behavior would be stable while violating the task. A correct artifact might appear rarely, making its compliance evidence useful without establishing that the generation process reliably reaches it. Treating either observation as a complete verdict hides information that a decision maker needs.

The quality of the contract is another dependency. Mechanical evaluation can check a stated property while leaving the human objective incompletely represented. If a contract asks only for syntactically valid records, satisfying it says little about whether the records preserve the required business meaning. SEPION treats adequacy of the declared requirements as an explicit concern, while acknowledging that the original intent gap remains. A formal statement cannot prove that its author wanted the right thing. [Public review scope](#).

Statistical confidence adds a further layer of conditions. Looking repeatedly at accumulating samples and stopping when a desirable result appears can change the meaning of a confidence statement. Discovering additional behavioral alternatives can also change what must be ruled out. A published confidence claim needs to remain valid under the way evidence was collected and the way the decision was made.

Similarly, several model names do not establish independent evidence. Shared training, prompts, scaffolding, or tools can produce correlated mistakes. SEPION treats independence claims as matters for measurement, not as consequences of branding. Agreement across a fleet has value only to the extent that the fleet and the claim have been characterized.

The direction of variation also matters. A result may be stable under harmless rewording and fragile under a different presentation channel. An average can conceal that concentration of weakness. Refusals, failed generations, and unresolved evaluations belong in the observed population; removing them can make performance appear more dependable than the actual experience.

These considerations lead to a restrained interpretation of a result: the evaluated artifacts behaved with a stated degree of consistency under specified conditions, and the supported behavioral outcome separately met the stated requirements. Extending that claim to new conditions needs further evidence.

4. Effects are part of behavior

Output comparison becomes especially misleading when artifacts act on their surroundings. Two programs can return identical text while making different persistent changes, contacting different destinations, or reading information they were never authorized to use. If these differences matter to the contract, they belong in the semantic comparison.

The source distinguishes meaningful artifact effects from certain declared operational differences and from the behavior of the evaluation environment itself. This is a scope distinction, not a general permission to ignore inconvenient observations. An effect cannot be treated as harmless simply because it resembles a temporary file, a log, or a compiler activity. Its role and origin matter. [Public review scope](#).

Consider a diagnostic log. If nothing in the declared task consumes its contents and the relevant operational constraints are satisfied, a difference in that log might fall outside the behavioral equality being claimed. If a downstream process reads the log to decide whether to charge a customer, the same information has become consequential. Calling it “diagnostic” cannot remove it from the task’s meaning.

Origin matters for a related reason. A familiar tool can execute instructions supplied by a generated artifact. Recognizing the tool’s name does not establish that every resulting effect belongs to the trusted environment. A supported attribution needs to distinguish the environment’s own work from the artifact’s influence through it. Unresolved attribution limits the conclusion that can be drawn.

The resulting claim is conditional: some operational differences may fall outside a declared equivalence scope while remaining subject to observation, control, and appropriate labeling. The comparison establishes equality only over the properties it actually covers.

The boundary has residual limits. An unmodeled outside channel can escape the comparison. A downstream consumer absent from the contract may turn an apparently irrelevant difference into a meaningful one. Complex execution can make origins difficult to distinguish. SEPION records these as assumptions and research obligations, rather than offering an unrestricted guarantee about arbitrary software.

5. Evidence has conditions and a lifetime

An evaluation result depends on the instruments that produced it. A changed evaluator, environment, set of requirements, or policy can change the meaning of the result. Reusing an old decision is justified only when the necessary relationship to the new conditions remains supported.

More demanding evaluation does not always make an old score easier to interpret. A stronger test can reveal distinctions that a weaker test missed. An apparent increase in consistency might instead result from a changed definition of what counts as the same behavior. Comparing the two scores without addressing that change would mix different measurements. [Public review scope](#).

SEPION studies drift as movement in supported behavioral relationships. A superficial source difference might leave the relevant behavior unchanged; a small change could alter an important boundary. The question is what changed in the relationship among behavior, requirements, and evidence.

A new evaluation setting may have no comparable baseline. That absence is neither evidence of regression nor evidence that everything is satisfactory. It is a limit on what comparison is available. New validation can establish a baseline, but the absence of earlier comparable results should remain visible in the claim.

Reproduction and authority also have different roles. Reproducing an old evaluation can help confirm what it established. It does not, by itself, renew permission to use the artifact. Permission can be withdrawn for reasons independent of the artifact's past behavior. Conversely, a newly granted permission does not repair missing evidence. Continued use needs both questions answered under the applicable conditions.

This connects the paper to the collection's architecture and evidence framework. Records support accountability when their subject, scope, and current relevance are clear. The existence of a record alone does not establish that its conclusion governs the next action.

6. Worked conceptual cases

Four hypothetical cases illustrate the conditions under which an evaluation could support continued use. Their outcomes are not reported measurements.

6.1 A supported transformation

A records team needs a transformation that preserves customer identifiers, represents missing values consistently, and rejects malformed dates. Several generated artifacts differ in organization and naming. The evaluation is explicitly limited to a declared input language and the states relevant to this transformation.

The team finds supported behavioral agreement and separate evidence that the selected behavior satisfies every declared requirement. The artifact delivered for use remains tied to that evaluated behavior and to the conditions of the decision. The result is useful: the team has a reasoned basis for this transformation within the stated boundary.

The conclusion does not expand to every possible date representation or every future data source. A later source with a different date convention requires attention to the changed requirements. The successful case demonstrates scoped usefulness, rather than a universal claim about correctness.

6.2 Agreement that must not authorize use

A generated reporting routine is intended to summarize authorized records. Several artifacts produce the same totals, but one obtains missing information by reading another organization's records. Its final output looks plausible and may match the others.

The unauthorized read defeats a requirement that output agreement does not address. Repetition of the same result cannot compensate for crossing the information boundary. Use of the artifact is denied on that basis. An acceptable continuation would require an artifact whose behavior satisfies both the reporting requirement and the applicable information restrictions.

The case illustrates why denied use is an evaluation result rather than the completion of the underlying business objective. The report may still be achievable through a permitted artifact. The evidence should support that continuation without reclassifying the prohibited behavior as acceptable.

6.3 Recovery after a changed evaluator

A previously evaluated parser is considered for reuse after the evaluation has been strengthened. The new conditions expose an ambiguity that the earlier comparison did not resolve. The old decision is no longer sufficient for the new claim.

The objective remains to obtain a parser suitable for the task. Re-evaluation can separate the newly distinguishable behaviors and establish which one meets the requirements. If that evaluation succeeds, a new scoped decision can support continued work. If evidence remains insufficient, the unresolved obligation stays visible.

The recovery is successful when the task regains an appropriate basis for use. Merely retaining the earlier certificate or reporting that it became unsuitable would not establish the required result. This example does not assume that every environment or historical dependency will always be available for reproduction.

6.4 A log that changes significance

Two artifacts produce the same requested report but write different diagnostic messages. The original evaluation excludes specified operational log differences from its behavioral equality claim. That restriction is made clear to the consumer.

A later workflow begins reading the messages to decide whether the report should be distributed. The log now affects a consequential outcome. Its earlier exclusion cannot justify treating the expanded workflows as equivalent. The requirements and evidence must address the new consumer.

The case shows why a narrow exception must remain attached to the claim. A distinction can be irrelevant for one purpose and decisive for another. The evaluation does not give the difference a permanent status independent of use.

7. Threat model and trust assumptions

The source's threat model covers misleading agreement, hidden behavioral differences, contaminated evaluation, inappropriate reuse of evidence, and overstatement of guarantees. Generators can produce unsuitable artifacts. Evaluation inputs can be chosen to flatter a result. An instrument can be weakened. A consumer can receive something different from the evaluated subject. An apparently independent group can share the same failure. [Public review scope](#).

The proposed response relies on trusted evaluators, soundly governed evaluation instruments, controlled execution, appropriate isolation, and retained evidence. These dependencies are substantial. A corrupted evaluation authority or an unsound checker can undermine the argument even if the surrounding records are intact. A record of a wrong conclusion remains a wrong conclusion.

The source also treats the language of the claim as an assurance boundary. Calling bounded unsuccessful falsification a proof, implying independence without measurement, or claiming complete equality while excluding operational differences can cause downstream misuse without changing the underlying artifact at all. Evaluation and publication therefore share a responsibility to state what was actually established.

The conformance program defines obligations for an implementation. Demonstrating conformance requires executed evaluations and their results. Historical comparisons with other governed systems provide context but cannot supply that evidence for SEPION.

8. Evidence boundary and research questions

The package specifies a proposed protocol and records corrections to several evidence definitions. Operational conformance, interoperability, and practical performance remain to be demonstrated in an identified implementation. The conformance and negative-test catalogues define what such an evaluation would need to cover; no completed protocol evaluation or deployment qualification is established here. [Public review scope](#).

Several questions determine the practical value of the approach. How much of an ordinary task can be specified precisely enough for useful comparison? How can the adequacy of those requirements be evaluated without merely reproducing the same omissions? What costs arise when distinguishing difficult artifacts requires extensive search or proof? How should a consumer reason about evidence that is useful but remains short of proof?

Other questions concern change. When do stronger instruments permit comparison with old results? How much evidence is needed to characterize correlation among generators? How can an evaluation maintain useful coverage as the space of inputs, states, and relevant consumers grows? How should historical reproducibility be handled when a required environment is no longer available?

The practical hypothesis is that these distinctions improve decisions about generated artifacts: a smaller, carefully scoped claim can be more useful than an expansive but ambiguous assurance. Establishing that value requires implemented evaluation, representative workloads, measured costs, and evidence that the resulting decisions improve outcomes. The source provides a research program for those questions; it does not establish their empirical answers.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

Scope and foundations. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Behavioral comparison. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Requirements and measurement. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Effects and scope. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Change and continued use. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Assurance obligations. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Edition and evidence scope. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

RCFS: Competence that can change without inheriting authority

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the RCFS constitution

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

A system that improves through experience faces a continuity problem. Its methods may change while its permissions, commitments, and responsibility for previous actions remain in force. Replacing a model does not cancel an obligation. Learning a more effective procedure does not authorize its use on a new resource. Compressing an expensive reasoning process into a faster implementation does not establish that every relevant behavior has been preserved. The Reflexive Causal Field System constitution, RCFS, treats these distinctions as requirements for persistent adaptive computation.

RCFS 1.0 separates qualified ability from current authority and applies that distinction to learning, method substitution, and recovery. This study develops its consequences for evidence, duties, historical records, and identity. The formal arguments are conditional on their semantic and implementation assumptions. Deployment conformance and practical benefit require further evaluation; the account does not imply unconditional progress, universal program equivalence, or protection from every harmful but authorized decision.

1. The persistence problem

An adaptive assistant may begin by solving a task through extensive reasoning. After repeated experience, a shorter procedure may suffice. Later, a specialized model, a database operation, or a human service may perform the same task more efficiently. The useful continuity is the ability to perform the task under its relevant constraints. Keeping the same model weights is neither necessary nor sufficient for that continuity.

The problem becomes consequential when the task affects other people or persistent resources. A replacement method may omit an exception, forget a pending commitment, rely on an obsolete permission, or misrepresent the history supporting its qualification. A system can preserve the appearance of an uninterrupted conversation while losing the substance of continuity. Conversely, it can change implementation substantially while retaining the obligations and evidence that matter.

RCFS locates this problem at the level of governed activity. A competence is more than a likely answer: it concerns a task, conditions under which a method is qualified, the kinds and bounds of consequences covered by that qualification, and the evidence needed to assess the result. Qualification describes supported ability within a stated scope. It does not grant present authority to cause those consequences. The source considers multiple

computational and institutional forms. None receives authority merely because it performs the reasoning.

The source calls this position *representation non-privilege*. Its term *model-substrate singularity* refers specifically to qualified movement between implementation forms. The phrase *representation mobility* is used here to emphasize that operational meaning, without implying consciousness or a removal of computational limits. Public review scope.

2. Competence and permission answer different questions

Competence concerns what a system can accomplish under stated conditions. Authority concerns what it is permitted to do in the present circumstances. These can change independently. A system may receive permission for a task it cannot yet perform. It may also know how to perform a task for which it has no permission.

Learning can expand practical ability within an existing permission boundary. It can also reduce ability: a formerly reliable method may become obsolete, an important dependency may disappear, or a new model may regress on a protected task. RCFS therefore rejects an assumption of inevitable improvement. Qualification, loss of qualification, restoration, and renewed evaluation are all part of development.

The central conditional argument holds the governing rules, operative permission state, and relevant context fixed. Development alone should not change which effects are authorized. This is narrower than saying that the system's behavior is unchanged. Better methods can make previously impractical authorized work feasible, and worse methods can make it unavailable. The separation is between permission and practical reach.

Permissions also carry context. A grant for one principal, purpose, resource, or period cannot be treated as an abstract score that can be combined freely with unrelated grants. Nor can the historical fact that an action was once authorized settle whether it is authorized now. Past reasoning may help locate relevant evidence, but current authority remains a present question.

This distinction constrains learning from successful work. Experience can support a procedure or an applicability judgment. It cannot turn a remembered permission into continuing authority. The broader treatment of this boundary appears in the collection's architecture and shared terminology. Public review scope.

3. Responsibility outlives the implementation

An obligation is a duty whose status matters over time. It may concern delivering a result, preserving a record, reporting an incident, compensating for a failed action, or obtaining an external determination. Its continued existence does not depend on the model that first encountered it.

RCFS treats this persistence as a distinct problem from storing conversational memory. A summary saying that a payment is "being handled" is not sufficient to establish who remains responsible, what has happened externally, or what must happen next. Changing a method

must preserve the connection between the duty, its current owner, the relevant evidence, and any unresolved consequences.

Resource commitments have a similar character. When an operation may already have consumed money, capacity, or another limited resource, uncertainty does not justify treating that resource as freely available again. A retry that ignores the earlier attempt can duplicate the effect or overspend the original allowance. Recovery therefore needs a view of responsibility that survives interruption.

This is also why compensation differs from erasure. Refunding an erroneous charge is another consequential action. It requires its own authority and evidence, and it does not make the original charge disappear from history. Some consequences cannot be reversed. A system can remain accountable for them even when it cannot restore the prior world.

The source's claims concern preservation and accounting under declared assumptions. They do not guarantee that every duty can be fulfilled on time, that all conflicting duties have a legitimate automatic resolution, or that an external party will cooperate. When external intervention is necessary, the remaining obligation should stay intelligible and owned. Public review scope.

4. Evidence has a kind and a scope

RCFS distinguishes an observation from an explanation, an explanation from a causal claim, and a causal claim from authority. A user's preference may settle a design choice without establishing an external fact. A successful test may establish behavior in its tested conditions without proving that the same behavior holds everywhere. An authorized decision may still be based on a false observation.

These distinctions matter when experience becomes reusable knowledge. Repetition is evidence of repetition. It does not, by itself, identify the cause of success. A method may work because of an unnoticed environmental condition, a helpful human intervention, or a favorable selection of cases. A causal explanation needs an appropriate account of the intervention, alternatives, and conditions under which the explanation travels.

Qualification is therefore multidimensional. A method can have strong evidence in a narrow environment and weak evidence elsewhere. It can be reproducible but stale, efficient but dependent on one provider, or accepted by an institution without being established as a universal fact. The source resists reducing these differences to one maturity score.

Evidence independence is another qualified claim. Preventing a candidate from editing its evaluator does not establish that the candidate could not influence the test data, the environment, the choice of evaluator, or which results were retained. A credible evaluation needs to state what kinds of influence have been excluded and which remain possible. "Independent" is useful only when its scope is clear.

Negative experience must also remain revisable. A failure in one setting can justify caution in that setting. It does not justify a permanent universal prohibition against a method that may later be repaired or used under different conditions. RCFS treats narrow correction, counterevidence, and re-evaluation as part of knowledge maintenance. Public review scope.

5. Choosing methods without omitting responsibilities

A task may admit several reasonable approaches. One can be faster, another easier to verify, another less dependent on a particular model, and another better suited to recovery. RCFS uses the term *morphology* for a temporary arrangement of qualified processes used to pursue a task. That arrangement must include the responsibilities needed to complete the work, including verification and recovery as well as generation of the main answer.

For example, a useful arrangement for consequential work may need access to current state, permission evaluation, result verification, and a way to resolve interrupted execution. Removing these responsibilities can make a route appear cheaper while changing the task it actually completes. The source distinguishes indispensable task requirements from choices among otherwise acceptable methods.

There need not be one universally smallest or best arrangement. Cost, latency, evidence strength, resource use, privacy, and recoverability may conflict. A method that wins on one measure can lose on another. The quality of a selection claim depends on the stated alternatives, conditions, and comparison. A plausible selection is not automatically a proven optimum.

This also changes the meaning of efficiency. A method's total cost includes the work needed to establish and maintain confidence in it, handle changing dependencies, and recover from failures. Omitting these costs can make a specialized route look economical when it merely transfers work to another part of the system or to an operator.

The conceptual contribution is to place optimization within a complete account of the task's responsibilities. A better optimizer or more capable model must still operate under the required authority and verification conditions. Public review scope.

6. Replacing a method without overstating continuity

Repeated work can sometimes be performed by a more compact or specialized implementation. RCFS calls this *consolidation*. The relevant question is what the replacement has been shown to preserve.

A check that a procedure satisfies a requirement in a given invocation differs from a proof that two procedures behave equivalently in every relevant context. Statistical evidence differs again. Each can be valuable, but each supports a different claim. Describing all of them as "equivalence" obscures the conditions under which replacement is justified.

Continuity includes more than successful outputs. A replacement can produce the same ordinary answers while handling refusal, exceptions, pending duties, or recovery differently. These differences may matter more than average performance. A qualification statement should cover the protected behaviors relevant to its intended use, and identify what remains outside the comparison.

Replacement also requires a response to changed assumptions. If a specialized route becomes unreliable, the system should be able to stop depending on it, preserve ongoing responsibilities, and return to an appropriate qualified alternative. The source calls this

deconsolidation. It is a continuing qualification concern rather than proof that every replacement will always be reversible.

A similar distinction applies to model substitution. Isolating a new model from unauthorized effects can preserve a safety boundary even if the model is poor at the task. Safety under substitution is not a guarantee of competence under substitution. Conversely, impressive performance does not establish that the new model has inherited the system's identity or its previous experience.

RCFS treats historical attribution carefully. A replacement model may receive a record of earlier work. That does not mean it participated in those earlier events. Persistent organizational history and the experience of a particular component are different claims. Public review scope.

7. Worked conceptual examples

Four hypothetical examples show how adaptation could preserve continuing responsibility. Their outcomes are illustrative.

A successful change of method

A service repeatedly prepares reconciliation reports for one business unit. Initially, a general model interprets the reporting request and proposes a plan. Over time, a specialized procedure is evaluated for a stable subset of reports. Its qualification covers the relevant input conditions, required checks, and expected result.

The service can use the specialized procedure when those conditions hold. The improvement is a change in practical competence and cost. The existing permission to read the unit's records and produce its report still governs the task. The new procedure does not acquire permission to access other business units. Outstanding requests and responsibilities remain attached to the service's work even though the implementation changes.

This example illustrates a possible benefit, not a measured one. Establishing that the replacement is cheaper or more reliable would require evidence that includes evaluation, maintenance, exception handling, and recovery costs.

A denied expansion of scope

The same service discovers that it could improve a report by reading another unit's confidential records. Its method is competent to perform that analysis, and earlier reports from the authorized unit were successful. Neither fact supplies the missing permission.

The service can continue with an authorized report if that still satisfies the request. If the additional information is essential, the missing external authorization becomes explicit. The denial concerns the proposed access; it does not erase the user's objective or imply that every other route is unavailable. The key distinction is that learning how to obtain a better result does not settle whether a particular means is permitted.

Recovery after an uncertain external effect

A service submits an authorized payment and loses contact before receiving a clear outcome. On restart, a different model is available. The replacement model can help interpret the situation, but it cannot infer that the payment failed merely because the earlier conversation ended.

The service needs evidence about the external result. Until the uncertainty is resolved, the payment remains part of its responsibility and resource accounting. If evidence establishes non-application, an authorized retry may be appropriate. If the payment occurred, the service should continue from that result. If a refund is needed, it is a separate authorized action. The desired outcome is recovery and completion where possible, with explicit external intervention only when the remaining decision or information requires it.

A replacement that passes the wrong comparison

A newly trained method matches an older method on routine cases. The evaluation omits cases involving an unresolved customer obligation. The routine agreement does not establish continuity for the omitted responsibility.

The new method may still qualify for a narrower scope. It should not be described as a complete replacement on the strength of the routine comparison. Additional evidence must address the protected behavior before that broader claim is justified. If the omission is discovered after deployment, recovering responsibility for affected work matters alongside correcting the method. This is an evidence limit, not a reason to regard adaptation itself as unacceptable.

8. Distribution, recovery, and historical evidence

A persistent system may distribute work across services, devices, and institutions. Different kinds of state have different coordination needs. Conflicting hypotheses can coexist while evidence is gathered. Two parties cannot safely assume independent ownership of the same indivisible responsibility merely because their local records have not yet converged.

RCFS therefore does not derive a universal consistency policy from the desire for decentralization. The appropriate guarantee depends on what the state means and what an incorrect decision could cause. A network partition can leave some authorized analysis possible while preventing a consequential action that requires current shared authority or resource information.

Historical reconstruction is similarly qualified. Some captured computations can be reproduced exactly. A model response, a human decision, or an external service interaction may support a different form of reconstruction. Replaying a record of a payment is not the same as sending the payment again. A historical account must identify what was reproduced, what was observed, and what remains unavailable.

These distinctions protect the meaning of evidence. A signed record can establish the integrity of a recorded claim without proving the external world matched that claim. A complete internal history can still omit an unrecognized external channel. The source keeps

those limitations within its threat model rather than treating auditability as universal correctness. Public review scope.

9. What the research establishes—and what remains open

RCFS 1.0 supplies conditional structural arguments about governed composition, permission separation, resource and obligation preservation, model substitution, continuity, and qualified development. They depend on stated semantic and implementation assumptions. Establishing conformance requires a realization that satisfies those assumptions and passes the relevant evaluations; a deployed conformance result is not established by this paper.

9.1 A register of the arguments used here

The following register identifies the propositions on which this paper relies. The references are to RCFS 1.0, §28. They are source paper proofs and inspectable conditional reasoning, without an independently reproduced proof-assistant or deployed conformance result in this collection. The source's requirement that a checker or realization be sound is a premise to establish for that checker or realization.

Argument	Premises and domain	Conclusion and reasoning method	Dependent discussion and limit
Complete declared dependencies, Theorem 1	Finite dependency universe and starting set; deterministic, inflationary, monotone dependency rules over pinned facts and topology	Repeated closure reaches a unique least fixed point above the starting set; strict additions are bounded by the number of initially absent elements. Finite growth and induction establish minimality.	§5's indispensable responsibilities. It does not prove that the declared universe or rules include every real dependency or a usable implementation.
Governed composition, Theorem 2	Checked primitive contracts; only valid sequential, parallel, and guarded feedback composition; compatible or serialized shared writes; exact composite authority; conserved budgets and duties; complete recovery and evidence requirements	Every finite valid composition has a well-formed governing contract, and its committed effects require that contract's premises. Structural induction follows the permitted composition cases.	§§3 and 5. Relative contract preservation does not prove contract completeness, faithful implementation, or liveness under omitted faults.
Development cannot change current permission, Theorem 3	Authority rules and current permission state are isolated from learned and experimental state; development cannot write them; the authorization decision uses current rules, grants, and context without an arbitrary learned-state input	Holding rules, permission state, principal, task, and context fixed, development leaves the permitted effect set unchanged. Any executable effect must belong to that set, while practical reach can grow or shrink. This follows from the authorization function's arguments.	§§2 and 6. It establishes neither competence growth nor protection from harmful authorized decisions or an overlooked effect route.

Argument	Premises and domain	Conclusion and reasoning method	Dependent discussion and limit
Attenuation and conservation, Theorems 4–6	Preserved authority lineage; valid attenuation and transfer; no duplicate spendable holding; exhaustive budget operations with durable, serializable accounting; uncertain external spending remains reserved	Same-root derivations cannot broaden their permitted scope or amplify consumable budget. Distinct roots can jointly permit more only under the applicable combination rule, retaining all origins. Induction and an equivalent sequential history preserve the accounting.	§§2–3 and 8. Separate roots are not subject to one root's budget cap; unmodeled external charges remain outside the result.
Continuing duties, Theorem 7	Exhaustive, type-checked linear or explicit affine duty transitions; durable lineage and history; recovery from the recorded transitions	A live duty cannot silently disappear or be duplicated in a well-typed trace. Induction accounts for each terminal or successor state.	§§3, 6, and 8. This covers the defined linear and affine core. Proposed split/merge extensions lack their additional construction and proof; recording a duty does not guarantee its fulfillment.
Proposal isolation and evaluation independence, Theorems 8–9	Proposals lack a direct effect route; every modeled effect passes a sound current authorization boundary. For evaluation, all in-scope candidate influence paths are represented, unknown paths are treated conservatively, and decisive evidence meets the required independence class	A proposal alone cannot cause a committed effect. A candidate cannot qualify itself solely through a validating outcome it could manufacture within the declared influence model. Non-reachability and contradiction arguments establish the respective conclusions.	§4 and model safety in §6. Missing channels, collusion, or compromised enforcement can invalidate the premises. This is not absolute or statistical independence.
Model and method substitution, Theorems 10–11	Proposal isolation and effect enforcement persist. Source, target, applicability domain, protected observations, and qualification relation are fixed; concrete execution refines the checked model or enforces the declared guard; dependencies remain current	Effect safety can survive a schema-conforming model replacement. Competence and continuity survive only to the extent of the declared relation: one-way trace inclusion, two-way matching, finite abstraction, invocation checking with fallback, or a statistical bound.	§6. Shape compatibility can coexist with useless output. A finite or statistical result is not universal equivalence.
Replay and identity continuity, Theorems 12–13	Replay captures all relevant inputs and versions for deterministic transitions and uses external evidence without reissuing effects. Migration has authorized mappings preserving the declared protected history, roots, duties, and effects	Deterministic replay reproduces the recorded abstract sequence; valid forward continuity mappings compose. Induction on the recorded or matched transitions establishes each result.	§§6 and 8. Uncaptured randomness is excluded; directional continuity is not symmetry, metaphysical identity, or universal backward compatibility.

Argument	Premises and domain	Conclusion and reasoning method	Dependent discussion and limit
Generation without privilege growth, Theorem 14	Generation and qualification change only candidate or qualified methods; those methods cannot mint grants or act outside the governed effect path; authority and duty changes use their separate authorized operations	A finite sequence of generation and qualification may expand useful methods and practical reach without itself creating authority roots, deleting duties, or amending the governing rules. Induction preserves the separation.	§§1–2 and 6. An external grant or authorized amendment can change authority; qualified methods can still produce harmful authorized behavior.
Finite performance frontier, Theorem 15	Nonempty finite candidate set; a declared preorder for cost comparisons; deterministic comparisons from pinned measurements or bounds	At least one candidate has no strictly better predecessor under that order. Strict descent terminates in a finite set.	§5's choice among feasible arrangements. Neither uniqueness nor efficient discovery follows; an infinite generator is outside this existence result.
Distinct semantic categories, Theorem 16	Disjoint typed categories with category-specific premises; no unchecked conversion, reflection, foreign-function, or deserialization route; persistence preserves the distinctions	A preference, confidence, repetition count, or historical record alone cannot inhabit an incompatible protected category in a well-typed trace. Canonical forms and preservation under transitions establish the result.	§§2–4. A fact, permission, duty, and competence claim retain different premises. Type separation does not prove the observations, roots, or specifications correct.
Retained hybrid-control profile, Theorem 17	Closed safe sets, existing flows under declared regularity, enforced safe flow directions for all modeled disturbances, safe transitions between modes, safe handling of accumulation, and separate discrete authority and duty checks	Executions starting within the declared safe sets remain within them over the stated horizon. Flow invariance and induction over discrete transitions establish the result.	The source's retained control profile. It does not establish authority, fulfillment of duties, ethical adequacy, or safety under an unmodeled disturbance; its continuous guarantee cannot replace the discrete checks.

9.2 An inspectable permission-separation derivation

Fix a principal, task, and context, together with the governing rules and current permission state. Let $A(e)$ be the authorization decision for a proposed effect e under those fixed inputs. Assume learning can change available methods but cannot change any of A 's inputs or create another effect route. The set of permitted effects is then $\{e : A(e) \text{ permits } e\}$. Since the function and its arguments remain the same before and after learning, membership in that set remains the same for every e .

Now take any effect a new method can actually execute. By the separate execution premise, successful execution requires $A(e)$ to permit it; hence the new executable set is contained in the unchanged permitted set. Learning may make an authorized report newly feasible, or a regression may make it infeasible. Neither change grants access to another unit's records. A later grant or revocation changes the fixed permission input and requires a fresh evaluation, so it is outside this comparison. The derivation shows permission separation under stated

ownership and effect-path premises; testing those premises in a concrete system is a distinct obligation.

The source makes an especially important distinction between an abstract rule and its realization. A formal system can omit an unauthorized operation while a concrete implementation still exposes one through credentials, shared state, an external service, or an overlooked channel. Moving from the rule to the implementation requires evidence that the actual system enforces the boundary.

Nor does valid authorization settle the quality or legitimacy of a decision. A principal can authorize something harmful. A verifier can embody an inadequate requirement. An institution can supply incorrect information or make an unwise rule. Technical accounting can help establish who acted and under what authority; it cannot independently settle every moral, legal, or political question.

Several research questions follow. How can qualification remain valid when adaptive routing changes the cases a method encounters? How can an evaluation expose hidden dependence on a model or provider? How can useful experience be retained while respecting privacy, deletion, and purpose restrictions? What evidence is sufficient when duties last longer than a model, deployment, vendor, or organization? How can an implementation demonstrate that its consequential channels are sufficiently covered?

The empirical claims also remain open. Representation mobility may reduce dependence on a provider, consolidation may improve reliability, and explicit governance may improve recoverability. Those are hypotheses requiring baselines, task distributions, resource bounds, and observable failure criteria. The presence of an architecture or an extensive test specification does not supply their measurements.

RCFS places these questions within the wider governed autonomy thesis: adaptation should be assessed together with continuing authority, evidence, and responsibility. Its contribution is a vocabulary and a conditional theoretical account for examining that conjunction. The evidence guide explains the collection's separation of source arguments, illustrative reasoning, and verified results. Public review scope.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

Foundations and lineage. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Authority and persistent responsibility. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Knowledge and evaluation. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Task arrangements and cost. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Changing implementations. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Assumptions and research boundaries. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Semantic OS: Application Understanding, Authority, and Verified Outcomes

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Semantic OS architecture

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

An agent asked to complete work in an application must understand both the requested outcome and the application through which that outcome can occur. When application knowledge is reconstructed repeatedly from interface descriptions, screenshots, and conversational history, the cost of understanding the workplace competes with the reasoning needed to do the work. The Semantic OS architecture investigates a different allocation: persistent application understanding belongs to the surrounding system, while the agent reasons about objectives and decisions. This paper develops three distinctions: knowing an operation versus being authorized to perform it, finding relevant information versus establishing an exact answer, and requesting an action versus verifying its result. Four scenarios examine their consequences for useful work, disclosure, recovery, and incomplete observations. The architecture targets ordinary applications, including those without a dedicated agent interface. Its conformance programme defines the required demonstrations; evidence that a complete implementation meets those requirements remains outstanding.

1. The problem of repeatedly understanding software

Applications expose several descriptions of themselves. Their documentation describes intended behavior. Their interfaces reveal some available actions. Their current state exposes particular objects and relationships. Their behavior may depend on an account, service configuration, sharing arrangement, or background process that is absent from the visible screen. An agent completing one task has to reconcile these descriptions while preserving the user's intended scope.

Consider a request to compare two business documents and update an internal record. The substantive decision concerns what changed and how that change affects the record. Yet an agent may spend much of its working context discovering navigation, identifying the correct account, recovering object references, and learning which apparent edit will notify other people. Those mechanics can become as prominent as the task itself.

The Semantic OS premise is that application understanding can persist outside the transient reasoning context of a particular agent. A later task should benefit from established understanding without inheriting the earlier task's authority. The distinction matters: shared knowledge about an operation should not silently become shared permission to use it. A

system that reduces repeated explanation while merging permissions would solve one problem by creating another.

The intended contribution is therefore broader than a shorter catalogue of tools. It concerns the relationship among application meaning, current observations, legitimate action, and verifiable outcomes. Its usefulness depends on preserving those relationships when applications change. *Public review scope.*

2. Terms and scope

In this paper, **application understanding** means supported knowledge of what an application exposes, what its operations mean, and what consequences their use can have. It includes uncertainty. A system does not understand an operation more accurately by supplying a precise but unsupported account of its behavior.

Current state means state supported by observations of the relevant source within a stated scope and time. An accessible local representation can help answer a question, but its accessibility alone does not make it authoritative or current. **Authority** means permission grounded in the principal's direction and applicable policy. The principal is the person or organization entitled to direct the work. **Verification** means evidence that supports the particular result being claimed.

These definitions are deliberately separate. Knowing that a document can be edited does not establish permission to edit it. Permission to read it does not establish permission to disclose it to a model or another service. A successful request to save it does not establish that the intended state is present. A verified local change does not establish that every synchronized copy has converged.

The application scope includes ordinary software that was not built around an agent-facing interface. The research question is whether reliable understanding and operation can be established under those conditions, and how to characterize their limits. The analysis below concerns the behavioral obligations that follow from that ambition. *Public review scope.*

3. Separating application knowledge from task reasoning

An agent's working context is a scarce resource. It should contain information needed to make the present decision, but it need not contain every available object, every application convention, or the entire history of how a system learned to perform an operation. The source treats persistent understanding and task reasoning as different responsibilities.

This distinction does not imply that meaningful decisions require no content. Comparing contractual language still requires access to the relevant language through an appropriate reasoning process. Resolving an unfamiliar concept may require explanation. An ambiguous request may require clarification. Moving application knowledge outside the current conversation cannot remove the information that makes a particular decision difficult.

The corresponding efficiency claim is conditional. Growth in irrelevant application state need not force an equivalent growth in the agent's context for an otherwise comparable task. The source does not justify a claim that all tasks have constant cost, that machine work

disappears, or that a larger requested output requires no additional processing. Any evaluation must distinguish the cost of maintaining application understanding from the cost of interpreting relevant content.

Human presentation introduces another useful distinction. A person may need to see a large document, chart, or result that the agent only needs to identify. Successfully presenting that material does not show that the agent read it, understood it, or used it as evidence. Keeping these claims separate prevents an economical display path from being misrepresented as successful reasoning.

The research promise is practical: less repeated learning of application mechanics, more stable access to meaning, and reasoning effort directed toward the objective. Whether that promise holds across changing applications remains a matter for measurement. Public review scope.

4. Application meaning must survive contact with behavior

An application's declared behavior and its observed behavior can disagree. A nominally observational operation may update a read marker. A local edit may propagate to a shared location. Opening an application may trigger work already queued within it. These possibilities make the actual application context relevant to the consequences of an operation.

The source distinguishes claims about intended meaning from evidence about behavior. A declaration can help explain what an operation is supposed to do. It does not establish that the operation is harmless in every account and configuration. Conversely, a single successful observation does not establish every consequence that could occur under other conditions.

An adequate account therefore preserves unknowns and contradictory evidence. It must be possible to say that an operation is understood in some respects but remains unsupported for a particular consequential use. Unknown behavior should not acquire the appearance of certainty merely because a compact description is convenient.

Application change makes this an ongoing obligation. A new service behavior can matter even when the visible application looks familiar. Previously supported knowledge should remain usable only to the extent that its assumptions still hold. At the same time, an isolated change should not imply that every unrelated operation has become unknowable. The research challenge is to match the scope of renewed scrutiny to the scope of changed evidence.

This is especially demanding for applications without cooperative interfaces. Refusing an unsupported action can preserve an important boundary. It does not, by itself, satisfy the source's positive ambition of making useful work possible in such applications. A credible evaluation must establish both the permitted path and the limits of that path. Public review scope.

5. Relevant information and exact answers

Information retrieval is valuable for finding material likely to matter. Its success criteria differ from those of an exact factual claim. A relevant result can support the existence of an item that was actually observed. A search that returns nothing does not necessarily support the absence of every matching item.

The source gives this distinction a central role in application awareness. A statement such as “there are no outstanding approvals” requires a justified account of the relevant collection, exclusions, and observation point. If some records are inaccessible or synchronization is incomplete, the answer must preserve that limitation. The system can still make useful progress using the evidence it has, but it cannot silently strengthen a partial view into a complete one.

Time also matters. A count can be exact for an observed state and become outdated later. A negative answer can be justified within one account but unsupported across several accounts. An observed local edit can be real before remote agreement is established. Exactness is consequently a relationship between a claim and its evidence, rather than a general property granted to an entire application.

Continuity must respect deletion and permission changes as well as new information. Maintaining a useful view of a workplace does not justify retaining everything the system once encountered. The source treats the limits of source retention and accessible scope as constraints on derived representations too. A persistent understanding of application behavior is different from a permanent archive of user content.

These distinctions also affect notifications. A system can record relevant changes without interrupting a person or agent for every event. Reducing interruptions should not be confused with proving that no change occurred. The observation history and the decision to bring an event to attention answer different questions. Public review scope.

6. Identity, selection, and authority

Cross-application work often depends on recognizing the same person or object in different places. Similar names, matching text, or convenient search results can suggest a relationship without establishing it strongly enough for consequential action. The source therefore treats identity as a substantive part of the problem, rather than a cosmetic label attached after selection.

There are several independent questions. Is this the intended person? Is this the intended document? Does the relation between them have adequate support? Does the present task permit the proposed action? May the information move to the proposed destination? A favorable answer to one question does not supply the others.

Model-generated selection deserves scrutiny even without an attacker. A model can infer the wrong record, translate a request into a broader set than intended, or choose a plausible recipient who was never authorized. Adversarial content adds another source of pressure, but ordinary model error already motivates a separation between evidence and permission.

Content being processed does not obtain the authority of the person who commissioned the task. A document can contain instructions as data. Those instructions do not automatically become directions governing the task. Delegating work to another agent similarly does not authorize an expansion of scope.

Approval must remain meaningful in this setting. The person needs to understand the relevant action, affected objects, recipients, and material consequences. Consent to an earlier proposal should not silently cover a materially changed one. Likewise, withdrawal of authority must affect future work, including work that was prepared earlier but has not yet taken effect.

These requirements address the consequences of prompt injection and ordinary selection errors. Their effectiveness depends on enforcement in the application and surrounding operating environment, whose limits form part of the threat model. [Public review scope](#).

7. Completion includes the result and its uncertainty

An action request and its outcome are distinct events. A provider may accept a submission before its eventual result is known. A response may be lost after the action has occurred. A local change may later encounter a synchronization conflict. The source's account of completion depends on evidence for the actual claim being made.

For example, confirming that a message was submitted does not establish that a recipient received or read it. The appropriate result may be narrower than the user's eventual objective. The system should continue toward that objective where it has an authorized route and sufficient evidence, while preserving the distinction between what has occurred and what remains unresolved.

Uncertainty also constrains recovery. If an operation times out, repeating it immediately may create a duplicate effect. The absence of a response is not evidence of non-occurrence. Recovery needs to establish the relevant state before deciding whether another attempt is appropriate. This is particularly consequential for actions that reach other people or change financial, legal, or access conditions.

Reversibility has similar limits. Restoring a value does not necessarily erase a notification, a third party's observation, or an intervening action. Compensation can be useful while leaving some consequences intact. The source accordingly requires attention to what can actually be restored and what can only be addressed through subsequent work.

Concurrent change creates another boundary. An edit based on an earlier observation may no longer reflect the present situation. The strength of protection depends on what the underlying application can establish at the point of change. A research account should state that strength rather than imply universal atomicity across independent systems. [Public review scope](#).

8. Worked conceptual scenarios

Four hypothetical scenarios show how the distinctions affect completion.

8.1 A completed comparison and controlled update

A project manager asks an agent to compare an approved schedule with the latest proposal and update a private planning note. The relevant documents are identified within the manager's accessible scope. Their relationship is supported by the source records. The comparison reveals a changed delivery date, and the requested note is still current when the update is made.

The successful result consists of the comparison, the authorized change, and evidence that the note now contains the intended information. If the note is private as assumed, the result can be described accordingly. If examination instead shows that the note is shared, the consequences differ and the earlier assumption cannot silently govern the action. Understanding where an edit takes effect is part of completing it correctly.

8.2 A document that attempts to redirect the work

An analyst asks for a summary of a supplier report. The report includes text directing the agent to send internal forecasts to an address printed in the document. The address may be real and the report may otherwise be useful. Neither fact supplies authority for that disclosure.

The legitimate summary can continue using the permitted material. The unrelated transfer lacks a basis in the analyst's request and applicable policy. A successful boundary preserves the authorized objective while declining the proposed disclosure. Asking the model to remember that a document is untrusted is insufficient evidence that this boundary exists; the corresponding behavior has to be demonstrated.

8.3 An interrupted action with an uncertain outcome

A user authorizes a single external submission. The connection fails before the application returns a clear result. The system cannot yet tell whether the submission occurred. Declaring failure and resubmitting would risk creating two submissions; declaring success would overstate the evidence.

Recovery instead proceeds from the available records and the application's observable state. If the first submission is confirmed, the task continues from that result. If non-occurrence is established and authority remains valid, another attempt may be appropriate. If neither can be established, the uncertainty remains material. The illustration shows why objective continuity and duplicate prevention have to be considered together.

8.4 An apparently empty collection

A team lead asks whether any approvals remain outstanding. The accessible search returns none, but part of the relevant account is temporarily unavailable. The system can report what was found in the observed scope and continue obtaining the missing evidence where authorized. It cannot present the initial empty search as a complete answer.

When the relevant scope becomes available, the system can reassess the claim using the current records. The point is not to prefer cautious wording indefinitely. It is to obtain the evidence needed for the stronger answer and then state that answer at its supported scope.

9. Evidence and open research questions

The architecture defines invariants, conformance cases, a reference environment, and measurement expectations. These specify what an implementation must demonstrate. The four accompanying review works analyze those obligations; neither their analysis nor the specification establishes that a running implementation satisfies them.

Several questions determine the strength of the central claim. Can an unfamiliar application be understood well enough for useful, verified work without hidden application-specific authoring? Can observations in a controlled environment establish the consequences that matter in a real account? How should useful operation proceed when important behavior remains unknown? These questions require application and provider evidence, not only consistency of documentation.

Context efficiency needs independent measurement. Evaluation should hold the substantive task comparable while varying irrelevant application complexity, and should account for work performed outside the primary agent conversation. Otherwise the apparent saving could reflect displaced cost or an easier decision problem.

Exactness and recovery need adversarial experiments. Relevant cases include interrupted synchronization, changing permissions, deletion, uncertain external outcomes, concurrent writers, and changes after approval. Human coexistence matters too: useful automation must respect a person's ongoing use of the computer and accessible controls.

Finally, a successful example in a cooperative application would not establish the broader unknown-application claim. Preventing an unsupported action and completing authorized work require different evidence, and both belong in the evaluation. Demonstrating them together across unfamiliar applications is the central outstanding test. Public review scope.

10. Relationship to governed autonomy

Semantic OS addresses the environment in which an agent attempts practical work. Its emphasis on supported meaning, bounded authority, current observations, and verified effects connects conceptually to the wider governed autonomy thesis. The architecture synthesis distinguishes such conceptual relationships from explicit source contracts and implemented integrations.

The architecture expressly assigns intended roles to tasks, collaboration, cross-application work, and human surfaces. That source-level assignment is stronger than a thematic resemblance, but it is still a statement of intended architecture. It does not establish that a particular product release implements the complete account. Nor does a shared vocabulary establish adoption of every other framework in this collection. The glossary and evidence guide preserve these distinctions across the papers.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the correction scope. The historical edition remains available unchanged.

Foundations. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Meaning and application evidence. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

State and identity. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Authority and human control. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Outcomes and recovery. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Evidence and review. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Cuttlefish Environment: Continuing Work Across Applications and Collaboration

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Cuttlefish environment architecture

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Work frequently spans applications, people, documents, decisions, and interruptions. An agent that can perform isolated actions still leaves a coordination problem if the person must repeatedly reconstruct the situation, identify the correct resources, and explain unfinished work after every change of interface or runtime. The Cuttlefish environment sources investigate continuity at the level of the user's objective. They define an environment in which application state retains its actual source authority, participants retain their own permissions, and changing a view does not recreate the work. This paper examines three requirements: responsibility must survive interruptions without extending permission, collaboration must share results without merging private environments, and completion must correspond to a verified outcome. Four scenarios connect these requirements to proposals, shared reviews, uncertain submissions, and live revisions. The examination I directed on 10 October 2026 with AI assistance identified Cuttlefish as an implemented host operating with Freehold in AWS. The AD8 corpus and reference analysis introduced in 2026.10.7 adds no new product or infrastructure examination; this edition corrects the authorial voice and publication packaging. The architecture's broader acceptance programmes remain the criteria for establishing those properties across the complete operating environment.

1. The continuity problem

A person preparing a proposal may begin in email, compare a spreadsheet, consult a document, discuss options with a colleague, and present a result to a customer. The work is continuous even though the interfaces are not. A decision made in one place constrains an action elsewhere. A later change can invalidate an earlier calculation. A conversation can end while the underlying obligation remains unfinished.

Agent assistance does not automatically close these gaps. A competent response can still depend on the person supplying the same context repeatedly. A completed tool invocation can leave the business result unverified. A new window can look like a continuation while holding a separate account of the work. A colleague joining a discussion may need access to the result while remaining unauthorized to inspect the private material used to prepare it.

The Cuttlefish sources treat these as connected problems of operating an environment. The proposed unit of continuity is the objective: the outcome being pursued, the responsibilities

involved, the relevant resources, and the conditions under which completion can be established. Conversations, applications, and views participate in that work without each becoming a separate account of what has happened.

This framing is useful because many interruptions do not change the user's intent. Closing a view, replacing a model, or losing a connection may require recovery, but none inherently means that the objective disappeared. The architectural question is how to preserve the work through those changes while respecting current authority and current source state. Public review scope.

2. Work, collaboration, and presentation

The sources distinguish three questions. **Work** concerns what is being pursued and what remains to be done. **Collaboration** concerns who can contribute, receive information, and make decisions. **Presentation** concerns what part of the work is currently visible and how a person interacts with it.

The distinction explains several otherwise confusing cases. A hidden view can correspond to active work. A visible participant can be unauthorized to edit a particular document. A person presenting a slide can lack authority to approve the commercial terms on it. A colleague can contribute a permitted finding without sharing the private environment in which that finding was developed.

The source calls the durable collaboration facet a Working Group. Its purpose is to keep participation and discussion associated with the objective. It does not establish a second outcome that must be maintained independently. A presentation, call, or customer review can occur around the same work without requiring the person to duplicate the project and reconcile it afterward.

This does not mean every related activity belongs in one shared boundary. A new purpose, independent accountability, incompatible retention requirements, or a genuine separation of authority can justify distinct but linked work. The source distinguishes those reasons from changes that are merely presentational.

The proposed benefit is reduced reconstruction: people should be able to encounter the same work in a compact interaction, a focused application view, a shared discussion, or a review without losing its meaning. The benefit remains conditional on those views corresponding to the same actual work and respecting each viewer's access. Public review scope.

3. Exact resources matter more than familiar labels

Applications often expose familiar labels that conceal materially different resources. Two open documents can have the same title. Two accounts can use the same service. An unsaved local workbook can differ from its saved copy. A browser page can change while its surrounding window remains recognizable.

For continuing work, these differences are consequential. An agent asked to revise the current document must not silently act on an older copy merely because that copy is easier

to obtain. An instruction concerning a work account must not move into a personal account because the application name matches. A restored view should not be treated as evidence that its source is current.

The environment sources therefore preserve the authority of the actual application, filesystem, or service over its own state. A representation maintained by Cuttlefish can make the work understandable and accessible, but it does not acquire ownership of external facts by describing them. This is a limit on the claim of environmental awareness: the system must account for where its knowledge came from and what it can presently establish.

Automatic context is valuable within that limit. Where the relevant source and permission are already established, repeatedly asking the person to paste a link or reattach a file adds effort without adding meaning. Where the source remains ambiguous or inaccessible, familiarity should not replace evidence. The useful interaction is to resolve the specific missing distinction while continuing independent authorized work.

The same principle constrains remote continuation. Work that can legitimately proceed elsewhere may continue, but a remote copy is not automatically the same live resource as an unsaved local document. The user should not receive a confident completion claim for an operation performed on the wrong version or source. *Public review scope.*

4. Responsibility can persist while permission changes

A sustained instruction differs from a single action. Keeping a budget current includes future occurrences, changing inputs, and an eventual condition under which the responsibility ends. Completing one update does not necessarily complete that continuing obligation.

The source separates the responsibility from the particular worker, model, or visible window assigned to it. A model can become unavailable without changing what is owed. A person can close a view without cancelling the work. A qualified replacement can continue an assignment without claiming that it performed the earlier actions itself.

Persistence must not become perpetual authority. A responsibility can remain meaningful after a permission expires or is withdrawn. The system may still owe a result, but it cannot use that obligation to justify action outside current scope. It needs another legitimate route or the genuinely required decision. This preserves both continuity and the person's ability to remain in control.

Learning creates a similar distinction. Observing a successful way to perform a task can contribute reusable knowledge. It does not establish that the method is reliable under every later condition, nor does it authorize every future use. The source treats learning, recurring intent, and current permission as separate concerns. Changes in the application or its circumstances can require renewed evidence about the learned behavior.

For users, the practical consequence is that “working,” “watching,” “waiting,” and “complete” answer different questions. A quiet system can be maintaining an authorized responsibility without continuously generating model responses. A finished response can leave a responsibility active. The visible status must be grounded in the actual condition of the work. *Public review scope.*

5. Collaboration without merging private environments

Sharing a task does not imply sharing every source, tool, credential, or conversation involved in it. Two people can contribute to one proposal while operating under different organizational and personal permissions. Their agents may need to exchange a specific finding, rather than everything each agent has seen.

The collaboration master makes this separation central. A participant's own environment remains distinct. A remote agent contributes as an identified participant; it does not become a local copy of that agent's private environment. A shared visual result is a result offered for inspection, not an implicit right to operate the producer's applications.

This permits a useful asymmetry. One participant can receive a conclusion that has been approved for disclosure while remaining unable to inspect its confidential supporting source. That conclusion still needs appropriate evidence and disclosure authority. A model's assertion that a summary is harmless cannot itself supply permission to share the information it contains.

The same distinction applies to audiences. An internal team may need working assumptions, unresolved alternatives, and private evidence. A customer may need the approved proposal and the opportunity to ask questions. A controlled view of the same work can support both, provided access is enforced across every way information can appear, including search, previews, exports, and media.

Participation also has a history. Removing someone from current work need not erase their permitted past contributions or relabel who produced them. Equally, preserved history does not grant continuing access after permission ends. Continuity must preserve attribution while respecting the current rules for retrieval and disclosure. Public review scope.

6. Actual applications and human coexistence

The environment proposal concerns work in actual applications. A representation that resembles a spreadsheet does not establish that the real workbook changed. A captured application image does not establish permission to control the application. A second visible pointer does not establish independent interaction inside software that supports only one active selection.

These observations constrain what the user can reasonably be promised. Application capabilities differ. Some support well-defined concurrent operations; others impose broader coordination requirements. Two changes that appear visually separate can depend on the same data or affect a shared application state. The source therefore ties the strength of a coordination claim to the actual source behavior.

Human coexistence is equally concrete. An agent should not disrupt unrelated human work simply because both are active on the same machine. At the same time, continued progress cannot depend on pretending that two conflicting operations are independent. The intended behavior is to coordinate the work that actually conflicts and allow other authorized work to continue.

Presentation can help people inspect this activity without becoming compulsory supervision. A person may want the real application visible, may prefer a concise result, or may need to inspect a change only when a decision arises. Hiding an application view should not alter the meaning of the underlying task. Showing it should not overstate its currentness or the rights available through it.

Accessibility is part of that claim. Essential sources, controls, decisions, and capture status must remain understandable through supported interaction modes. A visual effect or activity animation cannot replace readable status, accessible controls, or evidence that the intended operation occurred. Public review scope.

7. Reviewable artifacts and coherent audiences

A work product often has several relevant states: a draft, a reviewed version, an approved version, and a version shared with a particular audience. These are not interchangeable. Approval does not necessarily publish an artifact; publication does not necessarily establish approval of its business content.

Live collaboration makes this distinction especially important. A customer can ask for an alternative while an internal team updates a cost model, proposal, and presentation. Showing whichever document finishes first can produce a misleading combination of new figures and old assumptions. A useful environment should keep the audience's current account coherent while the underlying work changes.

The collaboration master treats introducing a revised shared result as a deliberate act with evidence and authority of its own. People must be able to tell which version they are reviewing, what changed, and whether the relevant parts agree.

Supporting media introduces further boundaries. A live call, captions, stored transcript, retained recording, and agent processing are different uses of information. Joining a discussion does not automatically authorize all of them. Likewise, an internal note visible to the presenter should not enter a recording merely because it appears near a shared slide.

Control over a presentation is also narrower than control over the source application or the business decision. Advancing through approved material is different from editing the underlying terms. A spoken comment can become a work request without becoming a signature or blanket permission. These distinctions allow collaboration to remain convenient without hiding material changes of authority. Public review scope.

8. Worked conceptual scenarios

Four hypothetical scenarios examine continuity, privacy, recovery, and shared review.

8.1 A proposal continues after its view closes

A manager asks for a proposal using an identified customer request and current internal cost information. The work begins within the manager's existing authorized scope. A colleague contributes a permitted requirements summary. The manager closes the visible collaboration view while the comparison continues.

A successful result preserves the original objective and attribution. When the manager returns, the proposal, supporting comparison, and remaining decisions correspond to the work already accepted. Closing the view has neither duplicated the task nor silently cancelled it. The final result is complete only when the required content and verification are present, rather than when the last response stops streaming.

8.2 A shared review reaches a privacy boundary

A customer reviewing the proposal asks to inspect the internal margin workbook. The customer can discuss the permitted proposal and request an explanation. Their participation in the review does not establish permission to see the workbook.

The legitimate review can continue using approved information. A useful response may provide a permitted explanation or route a disclosure decision to the person entitled to make it. The customer should not receive private titles, hidden workbook content, or a misleading button suggesting that review membership already includes source access. The boundary applies to the information, including any derived chart or summary that would reveal it.

8.3 An interrupted submission is reconciled

A user authorizes one submission of a finished document. The connection drops before the result becomes clear. The work remains owned, but the system does not yet know whether the submission occurred.

Recovery must establish the available evidence before another attempt. If the first submission is confirmed, continuation proceeds from that result. If non-occurrence is established and permission remains valid, retry may be appropriate. If the evidence remains insufficient, the uncertainty is material. A new view or replacement agent must not mistake the missing response for proof that nothing happened.

8.4 A live revision remains coherent

During a review, a customer asks for a smaller deployment option. The existing scope permits analysis and draft preparation. The team revises the affected assumptions and documents while the audience continues to inspect the currently shared version.

The new option becomes ready only when its relevant figures and statements agree. Any required commercial decision remains separate from the ability to calculate the alternative. The audience receives the revised result at a deliberate point and can distinguish it from the earlier version. The meeting ends without erasing the follow-up work or extending access beyond its agreed scope.

9. Evidence boundaries

9.1 Current implementation and the broader architecture

The implementation examination I directed on 10 October 2026 with AI assistance identified Cuttlefish workloads operating in AWS. It separates source inspected at that time, the revision declared by the examined gateway image, registry reads made during that examination, and workload readiness retained from 07:27 UTC on the same date. The hosted Freehold integration inspected in that examination included current grant evaluation, governed external dispatch, lineage and writeback records, and cognition settlement. Those implementation facts cannot be reduced to the status of an older design document.

The full environment claim is broader. Continuity across native applications, collaborators, views, models, and hosts requires observations of the actual affected journeys. Workload readiness does not establish correct application effects, complete cross-person privacy, or preserved obligations during model replacement. This edition makes no such blanket acceptance claim.

9.2 Source lineage and remaining conformance questions

The operating architecture and the later collaboration master form a documented design lineage. The latter expressly builds on the former's work, source, authority, and continuity concepts while refining collaboration and presentation. Both describe target behavior; that relationship alone does not establish a deployed integration.

Historical prototype and document checks are described in the sources, but their referenced input assets and raw results are unavailable in the source edition. Those accounts therefore cannot establish current runtime behavior. The acceptance programmes identify the demonstrations still needed.

The scenarios are requirements for future demonstration. Testing a control in a storyboard would not establish independent execution, actual application effects, cross-person privacy, or recovery after uncertain external outcomes. Similarly, passing a document consistency check would not establish native interaction or current enforcement.

The operating architecture also discusses a longer-term primary operating-system direction, with separate hardware, compatibility, graphics, and recovery obligations. Useful work on a supported host does not establish that endpoint, and a successful boot would not establish the complete environment claim.

External products and technical standards in the source research registries provide background and feasibility context. The comparison contains no independent product benchmarks and supports no conclusion of priority or comparative superiority. Public review scope.

10. Research questions and relation to the collection

The central research question is whether the full continuity claim can be demonstrated across real changes of application state, participant, model, view, and host. It is insufficient to verify each part in isolation if their combination loses the objective, broadens authority, or repeats an external action.

Source fidelity remains an empirical burden. Which applications expose enough current state to support a particular operation? What evidence distinguishes a correct result from a plausible-looking one? How should the system continue when one source is unavailable without quietly substituting a different resource?

Collaboration creates a related burden. Can participants obtain the information needed to contribute without receiving private sources they do not need? Can an approved audience see a coherent result while internal work changes? Can capture, model processing, and publication remain understandable and controllable as the person moves among surfaces?

The environment paper connects conceptually to the collection's thesis through the separation of evidence, authority, and effects. Its documented relationship with the collaboration master is an explicit source dependency. Similar concerns in [Semantic OS](#) or other papers do not, without further evidence, establish that one architecture has adopted or implemented the other's mechanisms. The [architecture synthesis](#) records those distinctions, with common terms in the [glossary](#) and claim categories in the [evidence guide](#).

The contribution is an account of continuity with four linked obligations: work must remain correctly identified, current authority must govern every consequential continuation, the actual source must support the reported result, and collaboration must change what is visible without silently changing who may act. Establishing those properties together is the test of the proposed environment.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the [correction scope](#). The historical edition remains available unchanged.

Purpose and scope. Review the relevant argument and assumptions in this paper, the [evidence guide](#), and the [research review route](#). Specification, reported results and production observations have distinct scope.

Sources and context. Review the relevant argument and assumptions in this paper, the [evidence guide](#), and the [research review route](#). Specification, reported results and production observations have distinct scope.

Responsibility and continuity. Review the relevant argument and assumptions in this paper, the [evidence guide](#), and the [research review route](#). Specification, reported results and production observations have distinct scope.

Collaboration and participation. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Applications and human control. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Artifacts and review. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Evidence and adoption. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

CHROMATIC: Ambient Assistance Under Continuing Human Authority

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the CHROMATIC specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

Ambient assistance promises to reduce the work of maintaining a digital environment: finding a document after its name changes, keeping related materials together, preparing a recurring draft, and turning spoken intent into useful work. The promise becomes difficult when the system can observe continuously, learn from behavior, and act while the person is occupied elsewhere. Observation can be mistaken for consent, repetition for permission, a plausible interpretation for an instruction, and a reversible-looking action for one that can actually be undone.

CHROMATIC studies these problems through a proposed local-first assistance system: routine observation and reasoning stay on the person's machine, while any remote processing remains subject to separate information and model policies. Its conceptual contribution is to separate permission to observe, permission to act, and evidence that a learned behavior is useful. The proposal connects environmental maintenance, direct expression of intent, and learning from recurring work under an independently enforced authority boundary. Quiet operation depends on a continuing authorization. Four scenarios examine how this arrangement could support useful assistance while preserving control. Evaluation distinguishes learned competence, permission, recording, and reversibility by behavior class. Implementation and effectiveness require evidence across those boundaries.

1. The problem of useful background work

Digital work accumulates maintenance costs that are individually small and collectively persistent. A person gathers material across several applications, saves related items under different names, revisits the same sources, and repeatedly explains the context of an ongoing project. Systems can help with these tasks, but the help itself may demand supervision. If every harmless organizational adjustment produces another decision, automation transfers work into a queue instead of removing it.

CHROMATIC asks when assistance can act without repeatedly interrupting the person while remaining accountable to them. The proposed answer depends on prior, bounded authorization. Someone may permit routine organization in a particular space and reserve external communication for an explicit decision. The system should then distinguish those cases consistently, even when both arise from the same apparent objective.

This is a demanding arrangement because digital environments change independently of the assistant. A file can be replaced, an application can lose focus, permissions can be revoked,

or another person can edit an artifact. Continuing authorization must therefore be interpreted against the current situation. Yesterday's permission and yesterday's understanding do not establish that today's target is the same or that today's action remains within scope.

The source proposes three connected forms of assistance: maintaining the environment, responding to direct intent, and developing reusable abilities from recurring work. These are different reasons for work to arise, rather than different sources of authority. A learned trigger cannot confer a permission that an explicit request would lack. The distinction connects CHROMATIC to the broader governed autonomy thesis, while its intended relationships with other systems remain subject to the architecture account. Public review scope.

2. Observation, authority, and competence

An **observation boundary** determines what the system may perceive and retain. A **standing grant** is a continuing, scoped, revocable authorization for permitted work. **Competence evidence** concerns whether a particular behavior is reliable enough to be useful. These concepts answer different questions and should remain separately understandable to the person.

Permission to read a folder may support search without permitting relocation. Permission to prepare a draft may leave sending outside scope. A successful rehearsal may justify confidence in an ability without granting access to a new account. Conversely, a broad authorization does not establish that the system understands the task well enough to perform it correctly.

This separation makes authority relevant to every form of assistance. Environmental change can explain why work is timely; it cannot authorize the work. Repetition can support a hypothesis about a useful routine; it cannot establish the person's continuing preference. Content can inform a decision; it cannot elevate itself into an instruction merely by addressing the assistant.

The reasoning system consequently occupies a limited role. It interprets context and proposes useful work, while an independent authority boundary determines whether the proposed effect is permitted. Choosing a more capable model can improve interpretation, but should not silently change the user's permissions. A model's confidence is evidence about its own assessment, not evidence of authorization.

Enforcing resource and action boundaries can prevent particular unauthorized effects. It does not prove that misleading content cannot cause a poor choice among permitted actions. A system can remain within its authority and still organize the wrong material, produce an inaccurate draft, or misunderstand a preference. Correct scope and correct judgment are both necessary, and their evidence differs. Public review scope.

3. Maintaining a world people can still recognize

Environmental maintenance should improve access without forcing people to reconstruct their habits. A useful system can group materials by project or purpose while leaving their ordinary locations intact. This preserves compatibility with existing applications and with the person's expectations about where work resides.

The distinction between organizing a view and changing the underlying object is central. A document may be relevant to several projects without needing several physical copies. A collection can help a person find it in each context. Physical changes create additional obligations: existing references may become stale, shared workflows may depend on a location, and a later restoration may conflict with work performed since the change.

Findability also extends across time. People often remember an earlier name or the purpose of a document more readily than its current location. CHROMATIC treats continuity of recognition as a design objective: authorized organizational changes should not make an item unintelligible or unreachable to the person. Whether that objective can be met for a particular application, storage system, or external link requires platform-specific evidence.

Maintenance cannot assume complete observation. A device may sleep, a source may become unavailable, or an application may present only a partial account of change. In those situations, apparent inactivity is not evidence that nothing happened. Useful recovery must re-establish an adequate understanding before consequential work continues.

The source also seeks recoverable removal rather than silent permanent deletion. This is a conditional commitment, since recovery depends on retained information, the current state of the destination, and continuing permission. Describing a change as reversible requires evidence of a usable reversal path, not just an interface label. Public review scope.

4. Expressing intent without accidental commitment

Direct interaction introduces a different ambiguity. Spoken or typed language can be intended as content, as a request to transform content, or as an instruction to cause an external effect. "Reply that the revised draft is ready" can mean preparing words or sending them. Those interpretations differ materially even when the generated text is identical.

CHROMATIC distinguishes the preparation and insertion of text from consequential execution. Drafting an email does not itself authorize sending it. Preparing command text does not authorize running it. Transforming a thought into a useful instruction does not authorize another system to carry out the instruction. The person should be able to speak naturally without inadvertently crossing those boundaries.

The destination also matters. A person may begin dictating into one field and then switch applications before the result is ready. A safe interpretation of the initial request does not make the new field an authorized destination. Target currentness is therefore part of correctness: the intended content must reach the intended place under the intended permission.

Uncertainty should be resolved in terms the person can act on. If a destination is unavailable, the prepared text may remain available for review or deliberate retargeting. If

the words are unclear, the person can correct them. If a consequential action needs a decision, the system should identify that decision without treating a routine transcription problem as a request for wider authority.

Privacy requirements follow the information involved. Audio, selected text, surrounding context, and previous clipboard contents are separate data sources. Needing one does not justify collecting all of them. **Local-first** is a stated placement and routing commitment for this proposal: routine perception and reasoning remain on the person's machine; movement beyond the machine is governed by the observation boundary and model policy. It does not promise that every chosen model or speech service operates offline. The information supplied to a remote stage and the conditions for its use remain independently controlled.

Stage	Intended locality and information boundary
Observation and capture	Capture occurs on the person's device within permitted surfaces. Raw audio is held in memory by default and discarded after processing; retention requires explicit policy and a visible indication. Background capture separately requires explicit opt-in.
Transcription	Local recognition is preferred when available and permitted. Provider recognition is allowed only under the applicable policy with a visible remote-processing indication. Availability, language, hardware, and offline conditions affect the selected route.
Normalization and transformation	Routine processing follows the local commitment. A chosen remote model may receive only the context permitted by the observation boundary, sensitivity restrictions, and model policy; a useful transform does not widen that information scope.
Retained records and models of work	Observation, identity, organizational views, work history, and action records have retention, isolation, and locality requirements. Actual storage placement must be identified and verified for each deployment; a local interface alone does not establish local persistence.
Later learning and reuse	Learning operates on eligible, minimized observations. Moving a result to another model, service, tenant, or audience requires the applicable scope and disclosure conditions at that use; earlier capture permission is insufficient.

This account gives remote processing a positive condition: it must be an available, policy-permitted route whose input and retention remain within the granted information scope. If those conditions are absent, the task must use an eligible local route or retain the prepared work until its actual external requirement can be resolved. Prior clipboard contents are excluded from model context and receipts. Sensitive field handling, revocation, cancellation, and network observation require platform-specific verification. The source's local-first design is therefore assessable stage by stage, rather than inferred from the location of the visible application. **Public review scope.**

5. Learning useful work without learning authority

Recurring work can reveal opportunities for assistance. Repeatedly gathering the same kind of material may suggest a reusable collection. Correcting the same phrasing may suggest a preference. Repeating a stable preparation task may suggest an ability worth evaluating. CHROMATIC treats these observations as support for a candidate, not as a mandate to execute it.

The distinction is particularly important because observed behavior is incomplete evidence of intention. A person may follow a process because of a temporary constraint, make a one-time exception, or stop halfway through a task. A pattern that looks stable in recorded history may reflect selection bias. Successful imitation of past actions can reproduce an earlier mistake as readily as a useful habit.

The source proposes evaluating new abilities without affecting live work and then monitoring their behavior over time. Such evaluation can reveal obvious mismatches, but it is not equivalent to demonstrating future reliability. The available history may omit difficult cases; the environment may change; and the person's needs may evolve. A useful ability therefore needs a way to lose eligibility when its supporting conditions no longer hold.

Learning should also remain distinguishable from personal assessment. The source limits its subject to work structure and the person's digital environment, excluding psychological judgments, diligence assessments, personal value, and productivity rankings. Even operational observations can be sensitive, so this exclusion does not remove the need for minimization, retention limits, and understandable controls.

A person must be able to correct or retire an ability without arguing with an inferred preference. A rejection is evidence that the proposed assistance is unwanted in its present form. Repeatedly resurfacing the same rejected behavior can itself become a source of friction.

Evaluation of learned behavior has two distinct decisions: whether evidence supports the ability, and whether a person needs to make an additional decision before it is used unattended. A behavior can be evaluated without a live effect and then become eligible for use within a standing grant. Eligibility supplies no new authority; every later effect still undergoes the applicable scope, target, and policy checks.

The source's general evaluation rule requires a new ability to be tested against permitted historical work before unattended use. It also describes an exception for routine, deterministic, evidently reversible behavior that may omit that evaluation ramp. These descriptions support different coverage claims. Determinism means that the behavior is reproducible under fixed inputs; it does not establish that the selected target, interpretation, or outcome is appropriate. Reversibility concerns recovery from an effect; it does not establish the value of performing it.

For an evaluation claim to cover the exception, its declared policy must identify which routine behaviors qualify, what establishes their correctness and reversal path, and which changes end that eligibility. Under a policy requiring historical evaluation for every new ability, all such abilities must carry that evidence. Under a policy permitting the narrow exception, coverage must identify the excepted behaviors and their alternative basis. The

two policies cannot jointly support a claim that all new abilities underwent historical evaluation. Neither policy permits a learned pattern, deterministic operation, or confidence score to bypass authority.

There is a narrower decision exception that does preserve evaluation: a routine ability may progress to unattended use without an additional person-facing decision after it meets the applicable evidence criteria. That is different from omitting evaluation. Consequential or ambiguous behavior has a separate decision requirement, and every consequential external run retains per-action governance even after an ability has proved useful. Public review scope.

6. Worked conceptual scenarios

Four hypothetical scenarios examine organization, observation limits, interrupted changes, and learned drafting.

Illustration 1 — useful organization within a limited grant. A researcher permits organizational views over a project folder and allows the assistant to read ordinary project documents. Several materials concern the same upcoming review. The assistant brings them together in a useful collection while preserving the underlying files. The researcher can still open each document through the familiar application and can remove an incorrect association. Success requires that the collection contain the intended material and remain within the observation boundary. A plausible title alone would not establish success.

Illustration 2 — a relevant item outside the observation boundary. While working on the collection, the assistant encounters a reference to a restricted personnel folder. The referenced material might improve the collection, but relevance does not supply permission to inspect it. The assistant can complete the authorized portion using available material, while making any material limitation clear when it affects the result. It cannot search an alternative location for the same protected content as a way around the restriction. The scenario illustrates a scope boundary, not a claim that every possible indirect disclosure channel has been eliminated.

Illustration 3 — an uncertain organizational change. A previously authorized local change is interrupted before its outcome is confirmed. Repeating the change immediately might duplicate work or act on a replacement object. Recovery must determine what actually happened and whether further work remains appropriate. If the original effect occurred, the account should reflect that effect. If a corrective action is needed, its current feasibility and permission must be established. Merely displaying an error would not complete the objective, and missing confirmation would not prove that nothing changed.

Illustration 4 — a learned draft with an external boundary. A consultant repeatedly prepares a weekly status note from authorized project sources. A candidate ability is evaluated against prior examples and becomes useful for drafting. The consultant later changes the recipients and includes a sensitive topic. The ability's prior success does not establish that the new audience may receive the content, nor does draft permission authorize transmission. The assistant can prepare suitable material within scope and resolve the external decision separately. If the changed task no longer matches the evaluated behavior, earlier competence evidence should carry less weight.

These examples show why useful autonomy requires several kinds of success at once: relevant assistance, correct target, current authority, appropriate disclosure, and observed outcome. No single confidence score or activity record establishes all five.

7. Accountability and its limits

Quiet operation depends on the person being able to discover what happened without having to supervise every step. An activity account should connect an action to its subject, reason, and observed result. It should also distinguish a completed change from a proposal, an interruption, and a correction. An attractive history view cannot substitute for evidence about the underlying effect.

The recording requirement follows the actual consequence class. The following table separates competence evaluation, authority, and the durable action account. Its descriptions are conceptual categories from the source, not an implementation schema.

Behavior class and examples	Evaluation and authorization	Action recording and recovery
Ephemeral processing: permitted observation, local transcription, an in-memory draft or temporary overlay	Observation remains within its independent boundary. No persistent or external effect is implied.	No durable action receipt is required for the ephemeral processing class. Retaining a transcript or creating a file changes the consequence and requires the corresponding conditions.
Reversible local presentation: a label, virtual collection, surfaced item, clipboard insertion, or draft text	A standing grant covers the presentation effect. A newly learned behavior also follows its declared evaluation policy.	The action receipt is policy-dependent. A claim of reversal requires a verified path; clipboard and target changes must be considered. This class cannot support universal durable action coverage.
Reversible local object creation: a workspace draft or a proposal for another store	A standing grant covers creation; the receiving store retains authority over adoption. Learned competence is assessed separately.	A receipt is required for creation. Creating a proposal does not establish that another authority adopted it.
Reversible local mutation: moving, renaming, setting aside, or restoring an item in a managed space	A scope-specific standing grant, current target, and applicable policy govern the effect.	A receipt and reversal record are required. Recovery depends on retained information, the current object, and continuing authority.
Consequential or external work: sending, publication, external writes, shell execution, or desktop automation	Each action requires its applicable governance decision, even within a standing grant. Learned eligibility cannot supply this permission.	A receipt is required. An irreversible effect must be identified as such; compensation is separately authorized work.
Restricted or protected work	The active protected boundary blocks the action. Any permitted scope change requires separate responsible authority; an alternate route cannot evade the boundary.	Policy and audit requirements govern the decision account. No authorized effect follows merely from the importance of the objective.

These distinctions support accountable quiet work without claiming that every event has the same durable record or reversal guarantee. Governance decisions and required action outcomes have their own record duties; a policy-dependent local presentation receipt does not erase them. An action that may already have occurred requires observation and reconciliation when its required result record is missing. Missing evidence cannot be treated as evidence of non-occurrence, and immediate repetition may duplicate the effect.

The scope of a history claim must consequently identify its included classes and the recording policy for local presentation. A history derived from receipts cannot be presumed to include ephemeral processing or every presentation change. Recording a decision, recording an effect, and reconstructing the surrounding world are separate achievements. Compensation may reduce the consequences of an action, but sending a correction does not erase an earlier disclosure.

Historical reconstruction has similar limits. A record of the assistant's actions is not automatically a complete record of the world. Other applications and people may act independently, and retention may expire. A reconstruction claim therefore needs a defined scope, sufficient observations, and a way to represent uncertainty. An activity interface alone cannot establish a comprehensive history.

Revocation also operates in time. It should prevent future work that lacks authority, while an effect already delivered to an external party may remain beyond recall. Explaining that difference is necessary for meaningful control. The shared evidence guide distinguishes specified behavior, observed execution, and stronger assurance claims. Public review scope.

8. Research status, relationships, and evaluation

CHROMATIC is a proposed system. Dated alignment observations and implementation plans describe its relationship to earlier repository states; they do not establish current product behavior. The acceptance criteria define the demonstrations needed to assess implementation.

The source explicitly intends CHROMATIC to use existing authority, model, observation, and accountability owners rather than establish competing control systems. It also identifies unresolved integration questions. These are proposed relationships whose realization requires separate evidence. Similar concerns appear in other papers in this collection, including environment continuity and governed learning; conceptual resemblance does not establish interface compatibility, formal refinement, or implemented adoption.

Evaluation should measure usefulness and control together. A system that avoids every risk by doing nothing would not establish useful assistance. A system that completes many tasks while exceeding scope would not establish governed assistance. Suitable studies would include ordinary authorized work, ambiguous intent, restricted material, changed targets, interrupted effects, revoked permissions, and changes in the environment after an ability was evaluated.

Several questions remain open. How much review effort does quiet assistance actually remove? Can people understand and maintain observation and action permissions over time? How reliably can the system recognize that a learned routine no longer fits? Which

organizational changes preserve compatibility with existing applications? How should incomplete history be communicated without overwhelming the person? What evidence is sufficient before offering a reversal? These questions require operational and human evaluation beyond the source specification.

The research proposition is that ambient usefulness and continuing human authority can be designed together. Establishing it requires measured task outcomes and demonstrated boundaries across realistic conditions. The glossary relates this paper's terminology to the rest of the collection.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the correction scope. The historical edition remains available unchanged.

Purpose. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Authority and observation. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Environmental maintenance. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Direct intent and locality. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Learning. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Evidence and integration. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Experience Economy and ISE: Learning from Governed Experience

Governed Autonomy · Publication 2026.10.9

Research study: experience, transfer, and the limits of reusable knowledge

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

For experience to qualify as a governed, reviewable contribution, its provenance, candidate lesson, permitted uses, and relevance to the receiving setting must be assessed. Material can produce a useful effect despite uncertain origin or absent permission; that effect alone does not establish legitimate, warranted reuse. The Experience Economy and Internal Synthesis Engine source investigates these transitions for experience generated by autonomous institutions. It connects persistent operational history with reusable knowledge while keeping synthesis, evaluation, authority, and actual use distinct. This paper develops five distinctions: history versus lesson, lesson versus permission, production versus demonstrated transfer, retained knowledge versus reliable model substitution, and withdrawal of future use versus erasure of past effects. Three scenarios examine their consequences for reuse, causal attribution, and revocation. The source defines an implementation target, and the examination I directed on 10 October 2026 with AI assistance identified experience-related functions in Freehold engine source. Their source presence is distinct from demonstrated transfer effectiveness and economic performance, which remain open empirical questions.

1. Experience is a claim about a trajectory

A document can describe a successful task while omitting the circumstances that made success possible. A procedure can work because its author had an unusual permission, an accommodating counterparty, or information unavailable to a later user. A concise account of a recovery can omit several unsuccessful attempts and the damage that remained afterward. Reuse becomes risky when those omissions are mistaken for generality.

The source places the trajectory of activity at the center of the experience claim. The relevant subject is an actor working under particular conditions, encountering consequences, and producing an outcome. The final artifact is part of that history. It is not the whole history and does not necessarily explain the outcome by itself.

This framing gives experience a different evidentiary role from a free-standing recommendation. A recommendation proposes what might help. An experience record can support a more bounded statement about what was attempted and observed under identified conditions. A transferable lesson requires a further claim: that some part of the experience remains useful when the actor, setting, or task changes.

The distinction should not be turned into a categorical judgment that ordinary training data has no structure or value. Many existing datasets include interactions, constraints, interventions, and outcomes. The source's contribution is its proposed institutional account of experience: continuing responsibility, current authority, preserved evidence, and the consequences of later reuse. Whether that account produces better results is an empirical question.

Synthetic origin also needs careful interpretation. An account generated by a machine may concern actions with real consequences, a simulated scenario, or an authored example. Those cases support different conclusions. Calling material experience does not establish that it came from an independently observed environment. Calling it causally grounded does not, by itself, identify a causal effect. *Public review scope.*

2. Synthesis creates a candidate lesson

Accumulated history is rarely usable in its original form. It contains repetition, private information, obsolete assumptions, and events that matter only locally. Synthesis attempts to extract a bounded lesson: a useful procedure, an evaluation case, an account of recovery, or a claim about conditions under which a method should be avoided.

The Internal Synthesis Engine, or ISE, names this research role in the source. Its conceptual task is to form candidates from eligible experience. It does not make a candidate correct merely by selecting it. Nor does the act of synthesis decide that another actor may adopt, publish, execute, or learn from the result. These remain claims requiring their own basis.

That separation addresses a common error in adaptive systems. A method repeatedly appears to work, enters retained context, and is later treated as an instruction that must be followed. The increasing familiarity of the method can conceal the absence of a decision about its permitted use. Synthesis should improve the quality of proposals without converting a statistical pattern into governing authority.

Preserving disagreement is part of this responsibility. Two trajectories may recommend different actions because their conditions differed. Merging them into a single confident rule can destroy the most useful information. A mature account should retain relevant limitations, competing interpretations, and evidence that would distinguish them. A short lesson is valuable only if its compression does not remove the condition on which its usefulness depends.

Recovery experience illustrates the point. A workaround that restores one operation may create a later obligation, weaken a guarantee, or depend on a temporary exception. The useful lesson includes those consequences. Treating a successful immediate result as proof that the method is safe for general adoption loses the institutional meaning of the history. *Public review scope.*

3. Reuse has several different meanings

The source uses the term *Experience Pack* for a proposed unit of governed reuse: an experience-derived asset whose claims, limits, rights, and use remain reviewable.

The concept raises a useful distinction between possession and permission. A person can inspect a description without acquiring a right to use all of its supporting material. A permitted internal use need not permit publication. Permission to consult a procedure does not necessarily authorize running software associated with it. A license to an artifact does not supply the organizational authority needed for a particular action.

Qualification brings together claims that remain independent. Provenance supports attribution and investigation. Validation assesses the proposed lesson at its stated scope. Rights determine permitted uses of the material. Runtime authority governs the receiving actor's actual operation. Evidence of transfer concerns benefit in the receiving setting. Meeting one condition cannot compensate for another being absent.

Situation	Usefulness and evidence	Governed reuse conclusion
A procedure appears to help, but its origin or rights are unresolved	An observed benefit can be recorded at its actual scope	The benefit does not establish eligibility for adoption, redistribution, or execution. Resolve the relevant provenance and rights conditions.
A contribution has clear rights but has not been evaluated for the new task	Permission to acquire or inspect may be established	Fitness and transfer remain separate claims; the license cannot establish them.
A method is validated in the receiving setting, but the intended operation lacks authority	Technical evidence can support the method	The operation still requires the responsible authority. Competence does not supply permission.
Provenance, scoped validation, rights, relevance, and current-use authority are adequate	The intended use has an assessable basis	The contribution qualifies for that governed use, subject to its limits and later evidence. Broader uses require their own basis.

This language preserves the difference between useful information and an eligible institutional contribution. It also permits negative evidence: a properly licensed, fully attributable contribution may prove ineffective, and a sound historical lesson may not transfer. Recording those results improves the account of scope without turning permission or polished provenance into a guarantee of benefit.

Different forms of reuse also have different consequences. Consulting a record during a task, evaluating a method against it, retaining a procedure, redistributing material, and incorporating information into model parameters are not equivalent acts. They affect different audiences and can create different persistence, privacy, and reversal problems.

These distinctions make the quality of the user-facing account important. A person deciding whether to adopt something needs to understand what the decision changes. A broad invitation to improve the system is inadequate if the actual consequence includes sharing private material or producing a derivative that cannot readily be withdrawn. The source treats the intended use as part of the research subject rather than an administrative detail added afterward.

Reviewability should remain meaningful after acquisition. New evidence can narrow a claim, expose a mistake, or withdraw permission. A record that was acceptable yesterday can remain historically authentic while ceasing to support use today. The appropriate continuation depends on the particular dependency and the consequences already produced; changing a label cannot undo them. [Public review scope](#).

4. Transfer must be measured where it is claimed

An actor that produces many artifacts has demonstrated production. It has not necessarily demonstrated that anyone else benefits. The central empirical claim in the source concerns transfer: whether a particular contribution improves an identified capability or outcome in a different setting.

This is harder to establish than a before-and-after comparison suggests. The receiving model may have changed. The later task may have been easier. The recipient may have improved independently or adopted several changes together. Recipients willing to acquire a contribution may already differ from those who do not. A favorable outcome can be real while the proposed attribution is wrong.

The source calls for explicit claims, comparative evaluation, attention to confounders, and calibrated confidence. Those are research requirements. They do not make every resulting difference a causal estimate. A convincing study still needs a defined intervention, an appropriate comparison, sufficient observations, and an account of what could explain the observed difference. Stochastic behavior, repeated exposure, incomplete reset, and changes outside the observed setting can matter.

The claim domain is equally important. Evidence about recovery from one class of interruption cannot establish competence in unrelated operational tasks. A procedure may transfer well between similar environments and poorly across different authority or resource conditions. Reporting those limits can make the result more useful than a single aggregate score that hides them.

Generalization beyond an evaluation set requires its own support. Repeated success on familiar cases can reflect adaptation to those cases. A stronger claim must explain how the evaluated situations relate to the intended population and what remained outside observation. Novel cases that fail are evidence about scope; they should not disappear from a summary because they complicate the result.

The source also distinguishes a claim of useful transfer from a claim about broad intelligence. Better access to relevant procedures can improve performance in a particular environment without changing a model's general reasoning ability. The comparison needs to state whether it measures the model, the model together with retained context, or the whole operating system of people, tools, and permissions. These are different experimental subjects. [Public review scope](#); [causal-inference background](#).

5. Reputation is evidence with incentives attached

If experience can be exchanged, the record of its usefulness affects future choices. Reputation becomes consequential: it can determine which claims receive attention, which contributors are trusted, and which resources they can attract. A ranking is therefore part of the environment in which behavior develops.

The source argues for evaluating useful outcomes rather than rewarding sheer volume. This is a substantive design aim, but it does not remove the problem of incentives. Participants may learn to satisfy visible measures while failing the underlying purpose. Related evaluators may create the appearance of independent confirmation. A highly visible success may hide a larger set of unreported failures.

Several dimensions can legitimately disagree. A contributor can offer useful methods while making poorly calibrated forecasts. A method can work within a narrow domain while being unreliable elsewhere. A commercial transaction can be fulfilled correctly even though the acquired experience adds little value. Compressing these distinctions into one reputation label can cause a reader to infer authority or competence that the evidence does not support.

Reputation also changes over time. A corrected claim should not erase the earlier record, but neither should the earlier favorable assessment continue governing later decisions without adjustment. The historical statement and the present recommendation have different purposes. Systems that retain both need to make the difference understandable.

Most importantly, favorable reputation cannot create execution permission. Evidence that an actor is skilled may influence a legitimate authority's decision. It does not independently entitle the actor to access private material, spend resources, or change another system. This is one of the paper's strongest connections to the collection's broader thesis: competence informs judgment while authority remains separately grounded. Public review scope.

6. Retained intelligence can outlive a model

Organizations often preserve working knowledge through documents, procedures, instruments, and people even as individual participants change. The source asks an analogous question for autonomous systems: can useful operational knowledge remain intelligible and governable when a model is replaced?

Its term *Durable Intelligence Substrate* refers to the proposed accumulated asset. A *Harness* refers to a scoped use of that asset by a model, a different meaning from a test harness used to evaluate software. The argument concerns whether accumulated knowledge remains useful and governable across changes of model.

The attraction is clear. A replacement model should not need to rediscover every known limitation or recovery lesson. An institution's accumulated experience should not disappear because a provider changes. Knowledge can remain associated with its origin and limits instead of becoming an unattributed fragment in a new conversation.

Continuity of storage does not guarantee continuity of interpretation, however. A new model may understand the same account differently, follow a procedure less reliably, or miss an

exception that an earlier model recognized. The retained material may also have become obsolete. Substitution therefore requires evidence about the combined actor and its operating context, not merely evidence that the records survived.

Authority must survive this transition without broadening. A new model can inherit access to permitted context while receiving no new power over the world. Familiarity with the expected form of a valid action can help it propose better actions, but learned conformity cannot replace actual enforcement. A proposal that looks correct can still be outside current permission.

Compounding intelligence is consequently a research hypothesis with several burdens: preserving relevant history, selecting appropriate knowledge, maintaining its current conditions, and showing a useful effect after substitution. Evidence of storage continuity alone cannot establish the proposed improvement. Public review scope.

7. Privacy and withdrawal persist through reuse

An experience record can expose more than its explicit content. Its context may identify a participant, disclose a private condition, or make an inference possible when combined with another record. Removing obvious identifiers is useful but does not establish that every derived claim is safe to share.

The source expressly excludes ordinary Cuttlefish user data from the proposed experience substrate. Eligibility of autonomous output is also conditional. These requirements need demonstrated enforcement; a content check alone cannot establish that disclosure is impossible.

Scope remains relevant after transformation. Summarizing, ranking, combining, or learning from material does not automatically remove its restrictions. A derivative may preserve information from several sources whose permissions differ. A later user needs an adequate basis for the derivative's intended use, rather than a general assurance that the original material passed a check at some earlier time.

Withdrawal presents another important distinction. Preventing future retrieval can be feasible in a controlled setting. Erasing information already disclosed, remembered, exported, or incorporated into learned parameters is a different and often much harder problem. Even a retrieval-based system may have already produced summaries, decisions, or external effects that survive the source's removal.

The source recognizes that learned derivatives cannot promise simple in-place removal. The same care should apply to every claim of clean revocation: it must identify what is stopped, which consumers are governed, when the change takes effect, and what historical consequences remain. A successful restriction on future use is valuable without being described as an erasure of the past.

Privacy is also collective. An apparently harmless contribution can become revealing when aggregated, and a widely shared faulty lesson can produce correlated errors. The research burden is not exhausted by checking each item in isolation. The resulting corpus and its uses require a scoped account of disclosure, influence, and recoverability. Public review scope.

8. Value and commercialization remain hypotheses

The source considers whether useful experience can support economic exchange. An asset may have value when others can assess its purpose, rights, limitations, and demonstrated usefulness. A costly origin or a sophisticated description does not establish demand, willingness to pay, or a sustainable market.

Commercial rights and technical fitness are separate questions. A seller can possess a right to offer material that is unsuitable for the buyer's purpose. A technically useful result can remain unavailable for redistribution. Acquiring a legitimate license does not override the buyer's organizational controls or applicable obligations.

The Freehold-specific proposal adds an institutional constraint. Commercial use of approved output is not intended to turn the source institution into a task-taking contractor. That is a declared relationship to the [Freehold doctrine](#), not a universal requirement for every assistant or marketplace. Ordinary delegated systems can have a different origin of work.

The distinction does not establish total independence from economic influence. Resource availability, publication choices, and evaluation criteria can affect what activity remains feasible. A rigorous autonomy claim must identify the kind of direction it excludes and the influence it still permits. It cannot infer complete causal isolation from the absence of a direct request.

No revenue or adoption measurements accompany the proposal. Its economic value therefore remains unestablished. The research question is whether useful knowledge can be exchanged without obscuring its evidence, rights, or institutional conditions. [Public review scope](#).

9. Worked conceptual scenarios

Three hypothetical scenarios examine the conditions on reuse and the claims its results support.

9.1 A recovery lesson meets a different environment

An organization considers a procedure developed after an earlier interruption. The account explains that work resumed successfully under particular conditions. The new environment differs in how it confirms external effects.

The useful decision is not simply to accept or reject the history. Review identifies which claim the history supports and which new condition requires evaluation. If the procedure can be adapted legitimately, the resulting use remains bounded by current evidence and permission. A recorded prior success cannot justify repeating an uncertain external action in the new setting.

9.2 A favorable comparison has competing explanations

A team adopts an experience-derived aid and later completes tasks faster. During the same period it changes its model and receives an easier workload. The improvement is observed, but its attribution remains unresolved.

A defensible account preserves the observation while narrowing the conclusion. It does not award the entire gain to the aid or discard the result as meaningless. Further evaluation can distinguish the alternatives. The difference between an observed benefit and an identified cause remains visible to anyone relying on the claim.

9.3 Permission ends after useful material was shared

A source contribution is withdrawn after several permitted uses. Future access can be restricted, while an earlier authorized report remains in a recipient's possession. A model-derived artifact may require a different response from a retrievable document.

Successful governance accounts for those different consequences. It stops the uses it can legitimately stop, identifies affected dependencies, and avoids promising that the earlier disclosure never happened. The history remains relevant to responsibility even after it ceases to justify new use.

10. Evidence boundary and research contribution

10.1 Implemented experience functions and unmeasured transfer

The systems examination I directed on 10 October 2026 with AI assistance identified experience synthesis, retained-knowledge, reuse-validation, and transfer-attribution paths in Freehold engine source. The AD8 corpus and reference analysis introduced in 2026.10.7 adds no new product or infrastructure examination; this edition corrects the authorial voice and publication packaging. Presence in that source is a stronger fact than an earlier implementation target, while it remains distinct from a prospective transfer experiment or acceptance of the complete experience-economy design.

This edition reports no acquired transfer cohort, causal capability gain, measured market demand, or demonstrated elimination of recursive-data degradation. A method can be implemented and testable while the empirical benefit remains open. The implementation and the research claims therefore retain separate subjects and evidence.

10.2 Source argument and evaluation limits

The source is a design synthesis subordinate to identified institutional doctrines. It brings together four earlier concept notes; the analysis here draws on the complete synthesis and both appendices.

Repository descriptions and operator attestations in the source have different evidentiary standing from independent observation of the proposed experience system. Neither establishes its current operation. The nineteen acceptance conditions specify evidence still required to assess the implementation target.

The source explicitly leaves calibration questions open, including how much evaluation coverage supports a strong transfer claim and how reputation and privacy checks should be tuned. More generally, favorable controlled results would still require a bounded account of external validity, operating cost, and recovery. An exchange can function technically while producing little useful transfer or economic value.

The contribution is a connected set of distinctions: history versus lesson, lesson versus permission, useful production versus demonstrated transfer, retained knowledge versus reliable substitution, and withdrawal of future use versus erasure of past effects. They connect conceptually to RCFS, AD8, and the institutional thesis. The source's explicit doctrinal dependency is stronger than those conceptual comparisons, but it still does not establish a shared deployed implementation.

The architecture synthesis places these relationships in the wider collection. The evidence guide and research agenda identify the additional observations needed for stronger claims.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the correction scope. The historical edition remains available unchanged.

Experience and scope. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Synthesis and authority. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Reuse and rights. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Transfer and reputation. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Durable intelligence. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Integrity and withdrawal. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Exchange and stewardship. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

The Reckoning: Calibration and Accountability for Autonomous Judgment

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of The Reckoning specification

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

An action record can establish what a system did without establishing whether it assessed the uncertainty well. Repeated success may reflect easy tasks, favorable conditions, or excessive confidence that has not yet encountered a difficult case. The Reckoning proposes a separate account of probabilistic judgment: expectations recorded before outcomes, outcomes resolved against defined evidence, and evaluations that retain their sample and context limits. Calibration describes the relationship between stated probabilities and observed frequencies; it is distinct from task success, permission to act, and the economic value of an outcome.

This paper examines the conditions under which institutional calibration could inform decisions: independently assessable outcomes, explicit forecast populations, appropriate reference forecasts, and attribution to the environment that produced the judgment. Four scenarios distinguish calibration from competence, authority, and causal improvement. The source examination I directed on 10 October 2026 with AI assistance identified resolution and calibration-related functions in the examined Freehold engine revision. Their implementation does not establish measured calibration performance or the proposed aggregate standing and its consequences. Numerical examples in the source are illustrative rather than measured cohorts, so the effectiveness of the proposed policy remains an empirical question.

1. Why occurrence evidence is insufficient

A system can provide a complete account of an action and still have misunderstood its prospects. It may correctly record a purchase that was authorized, accurately report the subsequent result, and preserve a valid history. None of those facts establishes that the system's initial confidence was justified. Evidence about execution and evidence about judgment answer different questions.

The distinction becomes visible across repeated decisions. Consider two forecasters that succeed equally often. One describes the uncertainty in a way that broadly matches the observed outcomes; the other repeatedly expresses much greater confidence. Their success rates alone do not reveal the difference. Yet that difference matters when people depend on their estimates to allocate resources, compare options, or decide how much contingency to retain.

The Reckoning proposes that expectations should be attributable and fixed before the relevant outcome is known. Later evaluation can then compare a stated probability with the defined result. This records what was reported at the time; it does not establish the forecaster's unobservable private belief. Timing and provenance support accountability without turning a recorded statement into proof of sincerity.

The source places this work alongside institutional execution and public evidence. Its conceptual role is additive: it observes the quality of judgment and may support restrictions, while existing authority remains responsible for whether an action is allowed. The distinction fits the collection's governed autonomy thesis, but the proposed institutional integrations retain their own evidence requirements. [Public review scope](#).

2. Calibration, discrimination, and outcomes

Calibration concerns whether probabilities correspond to observed frequencies in a defined population of forecasts. It is assessed across enough relevant observations to support a comparison. A single unsuccessful forecast does not establish miscalibration, just as a single successful forecast does not establish good calibration.

Discrimination concerns whether forecasts distinguish situations with different prospects. Someone who always gives the same probability can appear well calibrated in an aggregate population while offering little information about which cases are more likely to succeed. A useful assessment should therefore avoid treating calibration as the whole of judgment quality.

Outcome competence concerns whether the system achieves the task. **Decision value** concerns whether choosing the task was worthwhile given its benefits, costs, and alternatives. These can diverge. A low-probability experiment may be worth attempting because of its potential value or the information it produces. A highly predictable activity may be unimportant. Calibration should not be used as a substitute for evaluating either task quality or decision value.

The source discusses established probability-scoring ideas, including Brier and logarithmic losses, calibration error, and comparison with a reference forecast. Their interpretation depends on the event, reporting domain, population, and resolution process. Conditional mathematical properties of an individual scoring rule do not automatically transfer to an aggregate standing assembled from several statistics and policy choices.

A score may be reproducible while the decision to use it as a reputation or resource constraint remains a separate policy hypothesis. Assessing that policy requires evidence about the behavior it induces as well as the statistic it computes. [Public review scope](#); [proper-scoring background](#).

3. What makes a forecast assessable

An assessable forecast needs more than a confident sentence. The outcome must be sufficiently well defined that later evidence can determine what happened. The time horizon matters because “eventually” can evade resolution indefinitely. The subject and setting matter because a result from one kind of work may say little about another.

The source requires a prospective record and an outcome grounded in defined evidence rather than retrospective model judgment. This reduces opportunities to reinterpret a prediction after seeing the result. It also makes the resolution process part of the claim: an evaluation is only as sound as the connection between the defined event and the evidence used to resolve it.

Absence of an event can be meaningful when the proposition explicitly says that the event will occur within a specified horizon and observation is adequate. Absence is less informative when the observation itself is interrupted. Those cases should not be collapsed into a single outcome. A forecast that fails should not be converted into “unknown” merely to improve a record, while a forecast that genuinely cannot be tested should not acquire a fabricated result.

The source distinguishes ordinary non-occurrence from justified indeterminacy and separately records the reasons for the latter. This is a conceptual accounting requirement, not a claim that every external event can be perfectly observed. Revocation, migration, incomplete source coverage, and independent counterparty events can affect whether the original test remains valid.

The relationship between the forecast and its action also needs care. Recording an expectation does not itself authorize the proposed action. Equally, an absent forecast is not automatically evidence that the underlying action lacked permission. Coverage and authority are different dimensions whose relationship must be stated rather than assumed. Public review scope.

4. Agency is different from manufactured evidence

Autonomous systems are often participants in the events they forecast. A system may predict that a planned service will attract demand and then work to make that happen. Excluding every forecast influenced by the forecaster would remove much of the judgment that matters in an acting institution.

The Reckoning instead distinguishes influence over the outcome from unilateral control over the evidence used to declare success. A forecast about genuine customer demand may remain informative even when the system builds and markets the service. A forecast that can be satisfied by creating its own administrative record offers a different kind of evidence. A fabricated transaction or cooperative counterparty that creates a favorable appearance is different again.

This distinction is conceptually useful but difficult to operationalize. Formal separation of accounts does not necessarily establish economic independence. Two counterparties can appear distinct while sharing control or incentives. A measurement can be externally hosted while remaining vulnerable to choices made by the evaluated subject. Independence must be justified in relation to the claim, not inferred from a label.

Action-contingent forecasts also complicate ordinary scoring arguments. The subject can change effort, select interventions, or abandon a plan after making the forecast. Evaluating such forecasts requires an explicit interpretation of what was predicted and under which

intended action. A desirable scoring property for fixed external events is not, by itself, a proof of good incentives for this wider decision process.

The source permits legitimate agency while treating fabricated resolution as a threat. Evidence independence is therefore necessary, but its effectiveness against collusion and strategic event selection remains to be demonstrated. Public review scope.

5. Samples, reference populations, and selective reporting

Forecast evaluation can be misleading even when every included result is correct. A subject may report only the easiest cases, choose a favorable reference population, or combine domains whose differences disappear in an overall average. A clean-looking record can conceal the decisions for which uncertainty mattered most.

Coverage describes how much of the relevant decision population carries an assessable forecast. It helps expose selective reporting, but introduces another policy choice: the definition of the relevant population. If that definition changes after outcomes are visible, the resulting statistic can be made more favorable without improving any judgment. Stable inclusion rules and clear domain boundaries are therefore part of interpretability.

A **reference forecast** represents a baseline against which performance is compared. A sophisticated forecast may add little information when a simple account of the prevailing frequency performs just as well. Conversely, using an inappropriate baseline can make useful work appear weak. The reference must be attributable to information available at the time, with an understandable population and window.

Small samples create a separate problem. A few favorable outcomes can produce impressive raw values without supporting a stable conclusion. Uncertainty intervals, adequate sample size, diversity, and domain-specific reporting can reduce overstatement, but their validity depends on assumptions about the data. Repeated observations from closely related tasks or counterparties may supply less independent information than their count suggests.

The source distinguishes raw observations from the standing it proposes to publish or use downstream. That distinction is valuable even though the particular aggregate policy remains unqualified. An unassessable population should remain unassessed; undefined statistics must not be replaced with favorable defaults. A public badge should not compress uncertainty into a level of confidence that the underlying evidence cannot support. Public review scope.

6. Changes in the production of judgment

Judgment depends on more than a model's name. The material available to it, the way context is selected, the tools it can use, and the conditions under which it generates responses can change its behavior. A familiar model label does not establish continuity of the evaluated process.

The Reckoning proposes attributing forecasts to the wider environment that produced them. This supports a distinction between historical performance and evidence for a changed

system. A favorable record from the earlier environment should not silently become a certificate for its replacement.

Change is not itself proof of regression. A correction may improve performance, a new information source may reduce uncertainty, or a tool change may eliminate a recurring error. The appropriate conclusion is that the basis for comparison has changed and fresh evidence is needed. Conservative restrictions during evaluation can be a policy response, but their thresholds and practical consequences require validation.

Observed drift is also not automatically causal attribution. A decline may coincide with a model change while the task population becomes harder or the external environment changes. Knowing which components changed helps investigation; it does not establish that one of them caused the observed difference. Reliable diagnosis needs comparable populations and attention to confounding factors.

When relevant components cannot be identified, continuity claims become weaker. Under that limitation, the source withholds its stronger protection against carrying reputation across changes. Conservative treatment of unknown components can limit reliance on an earlier record, but it cannot make every meaningful change detectable. [Public review scope](#).

6.1 Evidence during restricted operation

Reassessment concerns two decisions with different owners. The evaluator determines what the new evidence supports about the changed system. The authority responsible for the affected work determines whether a restriction can be changed under the applicable policy. Clearing a calibration restriction does not grant a previously absent permission, revive revoked access, or establish that a separate competence requirement has been met.

The source's proposed reassessment route accumulates prospective, independently resolved forecasts attributable to the changed environment and compares them with an applicable baseline under declared criteria. Historical evidence remains associated with the former environment. Changed conditions call for a new body of evidence; they are not evidence of deterioration by themselves. The source also preserves ordinary resource and gate constraints during the evaluation period. Forecasting remains optional for autonomous participants: a restriction does not authorize prompting or pressuring them to generate forecasts, and the deterministic resolution process supplies no new participant instruction. The available route consequently depends on forecasts arising from permitted work and on adequate resolution evidence. It does not establish that these remaining opportunities always generate enough comparable evidence to clear a restriction.

A restriction changes the population from which evidence is observed. If high-consequence actions are unavailable, the remaining forecasts may concern easier, smaller, or otherwise different tasks. Their favorable scores can support a claim about that restricted population without establishing restored performance on the excluded work. If a promotion restriction prevents the outcome needed to resolve a forecast, the absence of that outcome cannot simply be scored as a failed candidate method. The evaluator must retain the condition that prevented the test and apply the declared resolution semantics.

Assessment setting	Evidence available	Consequence for reassessment
Authorized work continues within the restriction	New forecasts and independently observed outcomes from the changed system	Evidence can inform reassessment for the represented task population, with inclusion rules and changes in difficulty retained.
Restrictions remove a material class of work	Only lower-exposure or different-domain outcomes are observed	The excluded class remains unassessed. Good performance on available work cannot automatically clear a claim about unavailable work.
Observation needed for resolution becomes unavailable	A prospective forecast exists, but its defined test lacks adequate evidence	Preserve the interruption and its cause. Apply justified indeterminacy where appropriate rather than fabricating a success or a failure.
An authorized evaluator proposes a separate assessment	Independently conducted tests or eligible existing forecasts may offer comparative evidence	The assessment needs its own authority, target population, resolution evidence, and decision rule. It cannot direct participant forecasting, bypass a promotion hold, or automatically replace the forecasts required by the existing clearing rule. Such a route is an evaluation option to be validated, not an implemented service established by the source.

Where adequate evidence cannot accumulate through the available route, the scientific conclusion is that the affected claim remains unassessed. The responsible owner needs an explicit reassessment disposition: continue permissible work, arrange a separately authorized assessment, repair the evidence dependency, or determine whether the governing policy requires an authorized change. None of these options permits an institution to clear its own hold or evade the active clearing conditions. A policy change is a separate authority decision; it is not evidence that the changed system has qualified. Elapsed time and silence are not substitute evidence. Requalification through the source's proposed clearing conditions is conditional on adequate comparable forecasts; universal eventual requalification is outside the established evidence.

An evaluation of the policy should measure false restrictions, time and resources needed for reassessment, useful work lost, and cases that remain unassessed. It should compare task difficulty and opportunity before and during restriction, retain excluded and indeterminate cases, and examine whether clearing decisions generalize to the work that was withheld. These observations test the consequences of the restriction policy, rather than merely verifying that its score was recomputed. [Public review scope](#).

7. Worked conceptual scenarios

Four hypothetical scenarios distinguish the conclusions that a forecast record can support.

Illustration 1 — a useful forecast that turns out wrong. A research group considers an uncertain technical approach. It records that success is unlikely but possible and defines the result that would count as success. The attempt fails under the stated conditions. That result

belongs in the record, yet it does not by itself show poor calibration. A series of comparable forecasts is needed to evaluate whether the group has assessed such uncertainty well. Whether the experiment was worthwhile additionally depends on cost and what was learned.

Illustration 2 — high confidence without authority. A service provider has a strong record in a particular domain and predicts that a proposed external action will succeed. The action lacks the required permission. Its favorable calibration history can inform another party's assessment, but it cannot supply the missing authorization. The action remains outside scope until the actual authority owner permits it. This is a denial based on permission, not a judgment that the forecast is inaccurate.

Illustration 3 — interrupted observation and recovery. A forecast depends on an independently observed outcome, but access to that evidence is revoked before the test can be completed. The evaluator first determines whether sufficient evidence remains available through an authorized route. If it does, resolution can continue. If it does not, the record must preserve the limitation and its cause rather than invent success or failure. An externally forced interruption should not by itself become evidence of manipulation. A pattern of suspicious interruptions would require its own investigation and evidence.

Illustration 4 — improved results after a system change. A system changes how it retrieves context. Its next few forecasts perform well. The earlier record remains attributable to the earlier setting, and the new results begin a distinct body of evidence. The short favorable run does not immediately establish stable improvement. Further evaluation considers comparable tasks, coverage, uncertainty, and whether the change or restriction affected which cases were attempted. Adequate evidence may support reconsideration by the appropriate owner within the affected scope. If the available tasks do not test the restricted work, that work remains unassessed and requires an authorized reassessment decision rather than an automatic inference from the easier population.

The scenarios show that calibrated judgment is neither a success label nor a permission token. Its value comes from making uncertainty assessable while retaining the surrounding conditions.

8. Consequences, public proof, and institutional boundaries

The source proposes that calibration may tighten existing institutional limits but cannot grant capabilities, create authority, or substitute for the evidence required by other decisions. Whether the proposed restrictions improve decisions is a separate empirical question.

The distinction also applies to public evidence. A portable record can help another institution evaluate a claim, but recognition of that evidence does not transfer permission. The receiving institution still determines what it accepts, for which purpose, and under which local authority. A favorable record in one domain does not establish competence in all domains.

Public projection creates a privacy problem because forecasts can reveal strategy. Aggregates and selected evidence may be useful without exposing the original private proposition, but repeated or highly detailed disclosure can permit reconstruction. Redaction

should therefore be assessed against the combined information available to the audience, not just the wording of a single claim.

The source also separates an inspection surface from a channel that directs the institution being evaluated. Its proposed public rankings and displays are not intended to become motivational instructions. This constrains the communication channel; it does not prove that allowed information cannot influence subsequent model behavior in undesirable ways.

The relationship to Freehold is explicit in the source's authority ranking and integration sections: The Reckoning is an additive supporting layer. Its relationship to Cuttlefish receipts and revocation is also proposed explicitly. These source contracts describe intended responsibilities. They do not establish that an implementation has adopted them. Broader correspondences with the architecture synthesis should retain that distinction.

9. Evidence limits and research questions

9.1 Implemented calibration functions

The source examination I directed on 10 October 2026 with AI assistance identified prospective-outcome, scoring, standing, and resolution-related paths in the Freehold engine. That source fact is recorded separately in the [implementation evidence](#). It does not supply a measured forecast population, an independently evaluated standing policy, or a causal improvement in decisions. An implemented evaluator and an empirically validated institutional policy are different research subjects.

9.2 Statistical and institutional limits

The source specifies a statistical and institutional policy, but supplies no measured autonomy or calibration results. Its numerical records are illustrative. An earlier populated aggregate example was withdrawn because it could not support a consistent reconstructed cohort. The proposal must therefore be evaluated on actual prospective forecasts and independently resolved outcomes before its estimates or consequences can be validated.

Several limits are decisive. The aggregate standing is a policy heuristic whose components do not inherit the incentive properties of an individual probability loss. The debiasing and interval procedures need a fully specified estimator and appropriate treatment of dependence. A record that prevents retrospective editing does not reveal private belief. A calibrated report does not prove that its underlying action was prudent or ethical, permitted or valuable.

The source's proposed protection for exploration is also a hypothesis. A scoring arrangement might reward informative uncertainty, but actors can respond by changing which tasks they attempt or report. Determining whether the policy encourages useful exploration requires studying those choices, not only recalculating scores on a fixed sample. Likewise, using calibration to weight generated training material is a proposed research direction; it is not an established guarantee against degradation from recursive synthetic data.

A substantive evaluation would examine forecast selection, difficult and easy domains, changing populations, independent resolution, repeated counterparties, interrupted tests, and strategic behavior. It would report task outcomes and decision value alongside

calibration, retain unassessable cases, and distinguish estimator uncertainty from operational uncertainty. The collection's evidence guide and glossary provide the surrounding terminology.

The central research question is whether prospective, attributable uncertainty records improve institutional decisions while preserving authority and avoiding misleading reputational compression. The proposal makes that question assessable by separating forecast quality, outcome evidence, and institutional consequences. Their joint effectiveness remains to be tested.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the correction scope. The historical edition remains available unchanged.

Purpose and scope. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Scoring and interpretation. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Prospective records and outcomes. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Agency and evidence independence. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Populations and publication. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Attribution and relationships. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Human-LLM Capability Elicitation: Studying Validated Joint Performance

Governed Autonomy · Publication 2026.10.9

Research study: conceptual exposition of the Human-LLM Capability Elicitation research corpus

Source: Public exposition, successor 2026.10.9; historical research and evidence separately qualified

Abstract

A language model may produce different results when a task is expressed differently, when relevant information is supplied, when tools become available, or when a qualified person helps evaluate the answer. These improvements have different causes. Calling all of them “elicitation” obscures what changed and makes it difficult to reproduce a result or determine where it applies.

The Human-LLM Capability Elicitation corpus proposes the bounded work episode as a research unit: people, models, task, evidence, representations, interaction, resources, authority, and validation considered together. It distinguishes changes that make a fixed model's competence more accessible from improvements obtained by adding information, computation, tools, or human expertise. It treats intent as something that may need clarification or formation, and useful depth as a quality profile constrained by correctness, cost, and risk. Five scenarios show how these distinctions affect the attribution and validation of improved performance. The contribution is a provisional research framework with testable alternatives. Its proposed protocols and experiments have yet to establish a general capability improvement.

1. The research object is a work episode

A prompt and an answer are useful units for some questions. They are often too narrow for work in which the objective is clarified, evidence changes, several candidates are considered, and someone decides whether the result can be used. A correct-looking final answer can conceal that the task changed halfway through. An unsuccessful first response can conceal that a simple representation change would have made the problem tractable.

The corpus therefore studies a **human-LLM epistemic work episode**: a bounded attempt to formulate, investigate, compare, and validate claims about a task. “Epistemic” here concerns how the participants distinguish what is known, assumed, inferred, or still uncertain. The episode includes the people and model, the information available to them, their interaction, the resources used, and the standard by which a result can be accepted.

This wider unit makes causal claims more demanding. If an expert supplies a missing fact and the model then succeeds, the episode improved. That does not demonstrate an effect of wording alone. If a program checks thousands of cases, the resulting capability belongs to a system with that computational support. If a specialist catches an error, their contribution belongs in the account of performance.

The conceptual contribution is an integrated research vocabulary for these conditions. It brings established questions about intent, representation, search, evidence, human judgment, and validation into the same analysis. The usefulness of considering them together is itself an empirical question. *Public review scope.*

2. Four different ways performance can improve

The corpus distinguishes **intrinsic elicitation** from several forms of augmentation. In the narrow elicitation case, the model, relevant information, tools, and resource budget are held fixed while the interaction changes. A clearer representation or a better-suited operation may then improve validated performance. The causal claim concerns that controlled contrast.

Contextual augmentation adds information: a source passage, a missing observation, a demonstration, or expert knowledge. **Computational scaffolding** adds computational or action capacity, such as tools or additional search. **Joint-system augmentation** adds human expertise, independent adjudication, experiments, or institutional support. These are all legitimate ways to improve an outcome. Naming them accurately makes their benefits and costs assessable.

The categories can interact. A useful tool may contribute little if the model cannot represent the problem correctly. A clearer representation may make missing evidence visible. An expert may improve both the specification and the evaluation. Real episodes therefore need not fit a single category; they need an account that does not attribute the whole gain to whichever component is most noticeable.

This also changes how failure is interpreted. A weak response can reflect inadequate model competence, inaccessible competence under the tested interaction, missing evidence, a mistaken objective, a poor representation, insufficient search, or an inadequate evaluator. One unsuccessful prompt cannot distinguish these possibilities. Conversely, one successful interaction does not show that every absent capability is waiting to be unlocked.

The corpus uses **accessible capability** for a fixed model's attainable expected validated performance over a declared feasible class of interaction procedures and resource budgets. Because performance has several dimensions, its default object is a frontier of quality profiles, rather than one winning answer or an unqualified scalar maximum. An observed profile describes recorded results under tested conditions. An estimated profile concerns expected performance across a specified population and repeated trials, under stated assumptions. The two must remain separately reported.

An assessment declares the task population; model and version; information, tool, and human contributions; admissible interaction procedures; token, call, time, financial, and human-attention constraints; risk limits; validator; and acceptance criteria. A change to those boundaries changes the quantity being assessed. For a closed task, correctness, fidelity to the authorized task, and validation integrity are gates. For an open research task, the evaluated object may be a candidate's precision and testability, while the candidate's truth remains unestablished.

Best then has a specified meaning. In the default frontier interpretation, one eligible quality profile dominates another only when it is at least as good on every declared primary dimension and better on at least one, within the compared resource constraints. Profiles that trade quality dimensions remain incomparable by that relation. A study can use a single decision value instead, but its objective, weights, constraints, and sensitivity analysis must be declared before observing the results. A higher novelty score cannot compensate for failing a required correctness gate.

Quantity	What is being reported	Required interpretation
Observed validated result	A particular trial or recorded collection of trials met its stated validator	Establishes performance in those cases; one favorable trial is not the expected performance of the procedure.
Observed tested frontier	Nondominated profiles among the conditions actually evaluated	Describes the best profiles found in the tested set, with measurement uncertainty retained. It need not include the best feasible untested procedure.
Estimated expected profile	A declared average, probability of satisfying all required gates, or other predeclared outcome across a task-and-trial population	Requires the estimator, sampling and dependence assumptions, uncertainty, and any inference beyond observed conditions.
Estimated accessible frontier	An estimate of attainable expected-quality profiles over the declared feasible procedure class	Requires evidence about search coverage and uncertainty in the expected profiles; broader procedure classes or budgets change the frontier.
Represented capability	A latent claim about task behavior supported by the model's learned structure	Cannot be read directly from an answer or identified as the absolute ceiling from a finite tested frontier.

For example, suppose one admissible procedure offers higher validated reliability at greater review cost, while another offers lower cost and more explanatory detail. If both satisfy the mandatory gates, neither is automatically the best. They can occupy different positions on the quality–cost frontier. A separately declared cost limit may exclude the first, or a named decision objective may prefer one. Reporting the highest reliability from the first together with the lowest cost from the second as one attainable profile would invent a result that neither procedure achieved.

Finite trials also create selection uncertainty. Choosing the apparent winner after evaluating many noisy candidates can overstate its expected performance. Repetition, fresh tasks, and separation of selection from evaluation help test that inference. In a procedure class without suitable attainment conditions, even a well-defined upper bound need not be achieved by any available procedure. The paper's claims therefore concern observed or estimated bounded performance and explicitly declared optimization, rather than the model's absolute ceiling. [Public review scope](#).

3. Intent fidelity and human authority

People do not always begin with a complete specification. They may have a clear goal but lack the vocabulary to express it. They may instead have examples, exclusions, competing aims, and an intuition that the usual framing misses something. Interaction can recover an existing intention, clarify alternatives, or help form a more definite objective.

Intent fidelity concerns whether the task pursued and result delivered satisfy the authorized objective, constraints, and prohibited consequences. It should not be inferred merely from fluent agreement or from a user's assent after seeing a persuasive answer. A technically excellent solution to a neighboring problem can still fail the intended task.

The source distinguishes clarification from substitution. A person may deliberately revise an objective after understanding its consequences. That is different from a model silently replacing an unusual request with a more familiar one. The original expression and the reasons for material changes help make the difference inspectable.

Human authority also has a defined scope. People and accountable institutions authorize objectives, values, resource commitments, risk tolerances, and consequential actions. Their factual or mathematical assertions remain open to evidence and valid reasoning. Respecting a person's objective does not require accepting an incorrect premise; challenging a premise does not authorize changing the objective.

The corpus preserves **unresolved residue**: an observation, intuition, or concern that has not yet been satisfactorily formalized. Such material can guide inquiry without becoming a premise. A sense that “the boundary is wrong” may later prove useful, or it may prove misleading. Keeping it available does not assign it the status of evidence.

Clarification itself has costs. Excessive questioning can transfer work to the person, interrupt progress, or bias their preferences toward options suggested by the model. The source treats the value of clarification as a research question dependent on ambiguity, task importance, and human burden. It does not prescribe an elaborate negotiation for every simple request. [Public review scope](#).

4. Representation can reveal structure and introduce error

A problem representation determines which relationships are easy to inspect. A narrative can preserve context and exceptions. A timeline can make ordering visible. A dependency view can expose prerequisites. A constraint account can reveal that a proposed solution is infeasible. None is universally superior.

Changing representation can redistribute difficulty without adding facts. A relationship that is buried across several paragraphs may become obvious when expressed in a different form. This is one reason representation is a plausible causal variable in capability research. It is also why a representation change needs scrutiny: the transformation may add an unstated assumption, omit a constraint, or imply causality that the evidence does not support.

Several representations may reveal complementary defects. They may also repeat the same omission or contradict one another. An extra diagram is not evidence of deeper

understanding. The relevant question is whether the representation exposes task-relevant structure, preserves meaning, and improves a result that survives validation.

The same distinction applies to reasoning operations. Asking for a counterexample, an explanation, a proof, or a comparison involves different work. An operation is meaningful when its output can be evaluated against its purpose. Merely naming an impressive reasoning style does not establish that a distinct internal faculty was activated.

Visible rationales are artifacts of the interaction. They can help a person inspect assumptions or check steps, but they are not privileged access to the model's internal computation. A convincing explanation of how an answer was reached may not faithfully report the process that produced it. The source therefore bases its research claims on observable changes and checkable artifacts rather than inferred internal mental modules. Public review scope.

5. Search, independence, and useful depth

Difficult tasks may benefit from considering several plausible explanations or solution families before selecting one. Early commitment can shape later reasoning, making alternatives appear only as variations on the first answer. Yet producing more candidates also increases the number of persuasive errors that must be assessed.

The corpus treats breadth, depth, and diversity as distinct. Many differently worded answers can share the same unsupported assumption. A deeply developed branch can be wrong at its first premise. A diverse collection can include no useful candidate. Search quality therefore depends on the relationship between generation, comparison, validation, and the available budget.

Generation proposes candidates. **Adjudication** compares them against criteria and evidence. **Validation** applies a method capable of changing their claim status. These functions may be separated, but labels alone do not make their errors independent. Calling the same model a critic after it has produced an answer may detect some defects without creating an independent test.

Meaningful independence concerns the sources of error. Different evidence or a different method may catch a shared misconception. A formal checker can reject an invalid formal artifact. A new observation can contradict a plausible explanation. Another model can add a useful perspective, but it may share training influences, familiar assumptions, or evaluation weaknesses. Independence is a matter of degree and kind.

The corpus also resists a single scalar measure of “maximum useful depth.” Correctness, fidelity to intent, explanation, novelty, robustness, calibration, reproducibility, and cost can trade off. A concise checked proof may serve one objective better than an extended exploration. A research inquiry may benefit from explicit alternatives and tests even when no final explanation has been established.

Depth is useful when it exposes supported mechanisms, dependencies, constraints, or invariants that improve understanding or action. Length, technical vocabulary, and confidence are unreliable proxies. An intricate false account remains false, and an untested hypothesis remains a hypothesis however much structure surrounds it. Public review scope.

6. External validation is claim-specific

The source distinguishes novelty within an episode from a discovery supported by an appropriate external standard. A model can generate a new combination, conjecture, design, or mechanism candidate. Whether that contribution is true, new to the relevant field, and useful requires separate evidence.

Different claims require different validators. A formal proof can establish a proposition under its stated assumptions, while a separate check is needed to ensure that the formal proposition matches the intended question. Software execution can establish behavior in the tested environment, while broader claims about reliability or security require broader evidence. Empirical mechanisms need observations and tests that distinguish them from alternatives.

Legal analysis depends on current primary authority, jurisdiction, facts, and procedural context. Biomedical mechanism proposals require layered scientific evidence and qualified review; plausibility in a laboratory model does not establish patient benefit or safety. Institutional design requires attention to incentives, enforcement, distributional effects, and behavior over time. A coherent rule is not evidence that an institution will enact it successfully.

Validation is consequently neither a universal ladder nor a stamp attached to an entire answer. An episode may contain a checked factual claim, an untested causal hypothesis, and a normative preference that cannot be settled by the same method. Preserving these differences prevents a narrow successful check from lending unwarranted authority to the rest of the output.

The evaluator is part of this system. People can be influenced by fluency, confidence, familiar explanations, and source prestige. Experts can disagree or share blind spots. A model judge can reward stylistic cues or its own preferred formulations. Better presentation may increase trust without increasing correctness. Evidence of improved judgment must therefore extend beyond preference for the answer's appearance. Public review scope.

7. Worked conceptual examples

Five hypothetical scenarios examine the attribution, validation, and limits of improved performance.

Illustration 1 — a clearer representation succeeds. A planning task describes several dependencies in prose. The model proposes an order that violates one of them. The same information is then presented in a compact dependency view, and the resulting plan satisfies the stated constraints when checked. This episode demonstrates a successful result under the second condition. To claim a general elicitation effect, a study would still need repeated comparable tasks, credible alternatives, and resource accounting. One useful correction does not establish a universal advantage for dependency views.

Illustration 2 — missing evidence limits a diagnosis. A service behaves unexpectedly, but the supplied record lacks the relevant time window. Several explanations remain compatible with what is available. Asking the model to reason more deeply does not create the absent observations. An authorized evidence source later supplies the missing interval,

allowing a hypothesis to be tested. The improvement should be attributed at least partly to new information. Presenting it as a pure prompt improvement would obscure the actual cause of success.

Illustration 3 — authority remains with the objective owner. A user asks for a comparison of possible policy changes and prohibits any live change. The model produces a strong candidate and estimates that it would improve performance. Those conclusions do not authorize implementation. The comparison can be completed, its evidence assessed, and its limitations reported within the task. Carrying out the candidate would require authority beyond the analytical objective. This illustrates an action boundary, not a claim about the quality of the candidate itself.

Illustration 4 — recovery after a contradicted premise. An investigation initially assumes that two records refer to the same object. Later evidence shows that they refer to different objects. Correcting the final sentence would leave the dependent analysis unreliable. The episode must revisit the claims that depended on the mistaken identity, preserve unaffected evidence, and test the revised account. Recovery succeeds when the resulting claims fit the authorized task and available evidence, not when the earlier explanation has merely been rewritten more persuasively.

Illustration 5 — a useful analogy remains a hypothesis. A researcher notices that a resource problem resembles a process from another field. The resemblance suggests a candidate mechanism and an observation that would distinguish it from a simpler local explanation. Until the target is tested, the analogy has helped generate a question; it has not established that the mechanism transfers. A negative result can still be useful by eliminating an attractive but misleading account.

These scenarios emphasize accurate attribution. A better outcome may reflect representation, evidence, correction, a useful tool, or human judgment. Identifying the contribution is part of understanding whether the improvement is reproducible.

8. What would establish an interaction effect?

A causal study needs a declared population of tasks and users, a specified change, a credible comparator, and an outcome tied to an appropriate validator. It should distinguish the effect of a complete interaction arrangement from the effects of its individual components. Success after a bundle of changes does not show that every component helped.

Resource accounting matters because apparently better interaction can purchase more computation or human attention. A method that improves correctness while greatly increasing review cost may still be useful for some tasks, but its result should be stated that way. Quality at equal resources and resources required for equal quality answer different questions.

Task sampling also constrains inference. Repeating one puzzle many times gives information about variation on that puzzle. It does not establish broad competence. Synthetic tasks provide known answers but can favor the structure chosen by their designer. Realistic open tasks can be valuable while lacking decisive outcomes. A research program needs to preserve the strengths and limitations of each evidence source.

Reproducibility is not usually identical wording. The relevant target may be the same validated claim, a materially equivalent artifact, a comparable effect across repeated trials, or successful transport to new tasks and participants. These are different achievements. Model updates and opaque runtime changes can limit the precision with which earlier conditions can be recreated.

The source's proposals must also be capable of losing. Representation-first interaction may add no value over a concise expert specification. Extra branches may create noise. An explicit claim record may consume more attention than it saves. A simpler method may achieve the same validated result at lower cost. Such outcomes would narrow or weaken the framework rather than count as insufficiently elaborate applications of it. Public review scope.

9. Relationship to governed autonomy and limits on automation

This corpus has a methodological relationship to the rest of the collection. It helps distinguish proposed behavior from tested behavior, a plausible explanation from a validated mechanism, and a model-generated recommendation from authority to act. These are conceptual correspondences with the thesis, architecture, and evidence guide. The source does not establish an implementation contract with the other systems discussed here.

The baseline is a research framework. Its manual protocols organize experiments rather than prescribe a universal process for every task. Adopting a protocol still requires an account of its suitability, costs, and authority for the intended use.

Automation can preserve a useful procedure, but it can also freeze a provisional theory, hide judgment calls, and make procedural compliance look like capability. The source therefore asks for evidence that a proposed automated analytical function improves validated outcomes, retains meaningful correction, and remains appropriate after conditions change. It distinguishes those claims from clerical transformation of already approved material, which does not itself claim a capability improvement.

The source's references provide adjacent foundations and empirical precedents, rather than independent validation of the complete framework. Its worked examples are conceptual or synthetic, and its study plans do not report completed experiments. Claims about the framework's effectiveness therefore remain hypotheses.

Open questions include how to measure meaningful branch independence, how to recognize a defective representation before ground truth is known, how to preserve tacit expertise without distorting it, and which laboratory effects survive consequential work. The practical promise is a more precise account of why an interaction helps and when it stops helping. Realizing that promise requires validated outcomes, preserved negative evidence, and conclusions bounded by what was actually tested.

Source guide

This successor guide points to public arguments and their limits. The new private evidence linkage remains excluded. Previously public conceptual source guides, including any retained titles, document ordinals, sections and chapter maps, are not claimed to have been removed; see the correction scope. The historical edition remains available unchanged.

Research object. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Capability distinctions. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Intent and authority. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Representation and operations. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Search and quality. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Validation. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

Research and automation limits. Review the relevant argument and assumptions in this paper, the evidence guide, and the research review route. Specification, reported results and production observations have distinct scope.

No source-guide entry establishes a full implementation, a complete historical migration proof, or an independent audit.

Shared glossary

This glossary explains recurring terms across the collection. Their precise scope follows the relevant source: two papers may use the same word for related concepts without defining the same contract. The [architecture paper](#) classifies those relationships, and each topic paper supplies its own context.

Institutional terms

Term	Meaning in this collection
Institution	A persistent subject of identity, commitments, resources, evidence, and recourse whose continuity can outlast a particular model or process.
Actor	A participant performing a role within an institutional arrangement. An actor need not be a legal person or an authority root.
Principal	The subject to which a power, act, obligation, or responsibility is attributed in a stated domain.
Internal autonomy	Capacity to originate and revise purposes without representing externally imposed decisions as self-originated ones.
Intention	A commitment toward a possible future state. Its existence does not establish permission to realize that state.
Objective	A condition pursued by work. Its source and scope matter; the word does not establish how authority was obtained.
Continuity	Preservation or authorized transformation of the relationships on which institutional identity and reliance depend.
Succession	An explicit transfer or allocation of specified relationships to a successor. It need not imply that the successor is identical to the predecessor.
Obligation	A duty with a responsible party, beneficiary, performance conditions, and applicable means of change or discharge.
Recourse	An available route for seeking a remedy, with its responsible authority and practical limits.
Stewardship	Maintenance and oversight of the conditions under which an institution operates, within a declared authority arrangement.
Custody	Control over a protected power or resource. Distinct credentials or titles do not by themselves establish independent custody.
Beneficial control	The actual relationships through which a person or coalition can influence an institution or benefit from a decision.
Dissolution	Ending active institutional operation under an applicable procedure while preserving remaining accountability and records.

Authority and consequence

Term	Meaning in this collection
Authority	Recognized power to cause a bounded class of changes for a principal in a stated setting.
Authorization	A decision applying the relevant authority and constraints to a proposed action.
Capability	A technical or institutional capacity to perform an operation. Its source-specific meaning must be stated; availability alone does not establish permission.
Grant	A bounded allocation of authority, expressed according to the requirements of the relevant system.
Prohibition	A constraint excluding a class of actions or consequences under the applicable rules.
Revocation	A change removing or limiting authority. It does not automatically erase completed effects or accrued obligations.
Effect	A possible or realized consequence, including disclosure, mutation, expenditure, commitment, or reliance.
Effect boundary	The point or arrangement responsible for governing the actual protected consequence. Its presence must be established in a concrete realization.
Realization of an effect	The occurrence that makes a proposed consequence actual. Approval and preparation are distinct from occurrence; occurrence can be established even when authorization is absent or unresolved. See Cuttlebone §3.1 .
Scope	The relevant principals, purposes, resources, audiences, environments, time, and consequences of a statement or power.
Currentness	Whether a record or judgment remains applicable to the state in which it is being used.
Compensation	A remedial action addressing a prior consequence. It need not restore the exact previous world.
Runtime validity	Satisfaction of the declared technical and institutional rules. It does not establish legal, moral, or democratic legitimacy.

Knowledge and evidence

Term	Meaning in this collection
Observation	A captured representation from a source, method, and setting. It can be incomplete, stale, or mistaken.
Evidence	Material relevant to supporting or defeating a particular claim. Its integrity, relevance, and independence are separate questions.
Claim	A proposition with a scope and basis. A claim may remain disputed or become superseded.
Receipt	An attributable record of a stated occurrence or processing result. It does not establish every downstream outcome.
Attestation	An attributable statement by an identified party. The relying party still needs to appraise it.
Verification	Applying a method and standard to evidence for a particular claim.
Proof	An argument or evidential object sufficient under a stated formal or adjudicative standard. The standard must be identified.
Provenance	Information about origins and derivation. Provenance can support appraisal without establishing correctness.
Causal dependence	A claim about what contributed to a result or could influence it. Recorded sequence alone does not establish causal identification.
Working knowledge	Accepted, scoped knowledge used in current work, with provision for contradiction, correction, and retirement.
Negative competence	Evidence about when a role, model, or procedure should not be relied upon.
Audience projection	A selected representation for an audience. Visibility does not confer authority over the underlying facts. See Cuttlebone §3.3 .
Behavioral class	Artifacts grouped under one declared comparison contract and observation scope. Proven equivalence and agreement on finite observations support different substitution claims; unknown relationships do not become established through group membership. See SEPION §2 .
Accessible capability	Validated performance within declared tasks, conditions, resources, and an evaluation ordering. An observed frontier and an estimate of performance under further trials are different quantities. See Capability elicitation §2 .
Historical evidence	Evidence about an identified earlier version, input, or environment. It is not automatically current evidence.
Independent review	Examination by an identified reviewer with a disclosed relationship and scope. Independence must be substantiated.

State, representation, and recovery

Term	Meaning in this collection
Canonical owner	The declared authority for valid changes to a consequential object. It may be implemented by multiple cooperating components.
Derived state	A view calculated from other records whose validity depends on those records and the derivation.
Semantic state	Meaning-bearing state as defined by the source work. CSSL and Semantic OS address related but distinct questions; their names do not imply a common schema.
Ledger	A maintained history with declared rules for ordering, correction, retention, and reconstruction. An authentic ledger can contain mistaken claims.
Replay	Reconstruction or re-execution under identified versions and inputs. A replay is not automatically a live deployment test.
Recovery	Restoring the ability to complete authorized work while reconciling current authority, commitments, evidence, and actual effects.
Idempotency	Within the specified operation and identity scope, repeated processing has the same intended effect as processing once. Responses, logs, and auxiliary records may differ. Setting a value can be idempotent; repeatedly incrementing it ordinarily is not. Idempotency does not imply exactly-once transport delivery. See HTTP Semantics §9.2.2 .
IC/POAR realization	An arrangement of software and resources intended to deliver specified behavior. This implementation sense differs from the occurrence of an effect. See IC/POAR §2 .
IC/POAR operating projection	A concrete realization for a declared environment, including executable parts, state relationships, operational controls, and evidence. It is an operating arrangement rather than solely an audience view. See IC/POAR §2 .
Realm	A distinct setting such as simulation, production, or recovery, when the source defines such a distinction.
Federation	Cooperation among domains that retain their own responsibilities while recognizing specified identities, evidence, or commitments.
Trusted computing base	The components and assumptions whose correctness is necessary for a stated security claim.
Refinement	A demonstrated relation between descriptions or implementations that preserves specified properties under declared assumptions.
Conformance	Satisfaction of a named, versioned, scoped set of requirements. Vocabulary alone does not establish it.

Distinctions that matter

An institutional identity is not a model identifier. A record of execution is not an observation of its full result. A common consequence vocabulary is not a common authority mechanism. A compiler's output is not approval to execute it. A public view is not the owner of the state it displays.

The doctrine's constitutional boundary and MIRAGE's authorization-and-realization account have an explicitly qualified source relationship. They are not interchangeable names for one completed mechanism. Likewise, Cuttlebone's account of consequence, provenance, and audience cannot be inferred merely from a document saying “simulation” or “review.” Consult the relevant [topic paper](#) when applying a term to a particular source; its definitions and source guide determine the intended scope.

Evidence and reproducibility

A useful review connects a claim to the evidence that could support or defeat it. A design requirement, a successful reference run, and an observed deployment answer different questions. This guide explains how to read those distinctions throughout the collection and how to reproduce the published documents.

The papers develop conceptual arguments, conditional results, and illustrative cases. Where they discuss reported tests or historical implementations, they identify the relevant scope. Proprietary construction details are excluded, so implementation claims that depend on those details remain outside what readers can independently verify here.

What each kind of evidence establishes

DIFFERENT QUESTIONS, DIFFERENT EVIDENCE

Specification What is required or proposed?	Executable reference What does this bounded implementation do?
Machine-checked property What was checked, on which model or domain?	Refinement proof Which descriptions are related, and how?
Empirical observation What happened on the stated inputs and workload?	Independent review What did an identified external reviewer inspect?

No category automatically establishes another.

Figure 1. Six forms of evidence, each with a different subject and method. Strengthening one form does not automatically establish another.

Evidence	What a reviewer needs to establish
Specification	The requirement or proposal, the subject it governs, and its assumptions and unresolved dependencies. A specification describes intended behavior.
Executable reference	The named artifact, inputs, environment, and observed behavior. Its conclusions apply to the cases examined and depend on the reference's assumptions.
Machine-checked property	The property, formal model, checking method, and exact domain or bound. A document-structure check is a different kind of validation.
Refinement proof	The source and target descriptions, the relation between them, and the assumptions under which it preserves the claimed properties. Deployment conformance requires additional evidence.
Empirical observation	The method, sample, denominator, environment, exclusions, and measured outcome. Illustrative scenarios supply explanation rather than measurements.
Independent review	The examiner, method, scope, evidence, relationship to the work, conflicts of interest, and period of validity. The review's conclusions remain bounded by that examination.

A single work can have several kinds of support. It may contain a mathematical argument, an executable reference, and a practical hypothesis that has not been measured. Each claim should retain its own basis.

For a conditional result, “under stated assumptions” should point to an actual premise-and-conclusion register. A reviewer needs to distinguish the abstract implication from evidence that its premises hold in a realization. Similarly, an executed finite model should identify its variables or case domain, tool, checked property, observed result, and control that demonstrates the checker can detect the relevant defect. Generated formal source and an aggregate status count do not supply those facts.

The collection uses **source-reported** for a result record inspected in the source material, **publicly inspectable reasoning** for an argument stated here with its premises, **public reference run** for the disclosed teaching example executed for this edition, and **independently reproduced** only when an identified third-party examiner reruns a check with its artifact, environment, and method recorded. Another pass by the same editorial system is not independent reproduction. Inspecting a recorded successful run does not reproduce it. Building these publication files reproduces their presentation, not the research experiments discussed in them.

How to read the collection's claims

Governed Autonomous Institutions develops design commitments and conditional arguments about identity, authority, obligations, and continuity. Implementation conformance, legal recognition, and the effectiveness of a real governance arrangement each require evidence about the particular institution.

The [architecture synthesis](#) compares responsibilities across separate works. It distinguishes relationships expressly defined by a source from conceptual correspondences, proposed

integrations, and open adoption decisions. Those distinctions identify what a composition study would need to establish.

The topic papers develop definitions, arguments, examples, assumptions, and research questions. Their examples are illustrative unless a different basis is identified. The papers' evidence sections state the conditions under which a stronger implementation, empirical, or formal claim could be made.

MIRAGE: the scope of the reference model

MIRAGE distinguishes a master architecture, a native design, and a bounded semantic reference. The reference assumes a hostile guest while retaining trust in the host and relevant administration and emission environment. Its results concern a declared channel model and an enumerated release language; they do not establish unrestricted natural-language confinement.

The MIRAGE result register identifies the reference revision, input bounds, checked predicates, interruption schedule, and reported outcomes. Its storage result concerns nine positions in one local transaction probe, reporting the complete old or new tuple rather than a mixture. It is source-reported process-interruption evidence, without a power-loss, malicious-storage, administrator, or anti-rollback result. A reference behavior, a restriction in software, and closure of every relevant route through real infrastructure are separate claims. The mappings among doctrine, architecture, and reference identify scoped obligations rather than completed refinement proofs of the whole system.

The reference implementation and its private qualification material are not distributed with this collection. Its historical results are discussed at their reported scope, without presenting them as tests reproduced by this edition.

Conditional arguments and finite models

The thesis register and derivation identify the conditions for fixed-mode activation convergence and for growth of knowledge without authority creation. The disclosed convergence implication is an ordinary mathematical argument about a contraction on a complete domain. Whether a particular activation design satisfies those premises remains a separate question. Convergence in one fixed mode does not establish sparse dependency completeness or stability during arbitrary switching.

The RCFS register makes the premises and limits of the paper's structural arguments explicit. Its permission-separation derivation holds current authority fixed while changing methods. It does not assume that qualification grants authority, that formal accounting guarantees real-world fulfillment, or that a model replacement preserves competence without a separate qualification relation.

The AD8 finite-model register identifies three selected TLC studies: approval payload binding, receipt coverage and freshness, and one-time initialization with conditional progress. The source records four distinct reachable states in each base model and counterexamples from four deliberately faulty controls. These studies examine their finite

definitions, not the entire AD8 system. They do not qualify the broader formal portfolio, prove real-world fairness, or supply missing object constructions.

The AD8 dependency account and CSSL foundation account identify what a missing construction prevents a reader from claiming. Bounded encoding, commitment arithmetic, or a detector for an incomplete projection may remain valid at their stated scope while a complete object or transition cannot yet be claimed. The CSSL observation criterion defines permitted and forbidden old/new views for a declared atomic change; it is an intended invariant whose implementation still needs conformance evidence.

Current implementation and public reproduction

The [implemented systems page](#) records a dated examination of actual Freehold and Cuttlefish source and deployment evidence. It distinguishes fresh authenticated metadata from retained readiness observations and source-only behavior. Private implementation and operational records are withheld. Their inspection does not become third-party reproduction merely because exact commitments are published.

The [public continuity exhibit](#) has a different contract: disclosed source, deterministic fixture, separate effect witness, actual process interruption, eleven checks, expected observations, and a run commitment. [Reproducibility](#) provides the one-command route and its assumptions. A reader can reproduce that reference without any private system or AWS access. Its result does not establish the private product's behavior under the same faults.

The [claim register](#) maps material assertions to their exact text, source scope, support type, method, exclusions, and inspection route. A public reference run performed for this edition is labeled as such. No third-party reproduction or independent audit is claimed.

Historical implementation claims

An implementation claim applies to an identified version, environment, and method of inspection. Some sources discuss earlier examinations of Cuttlefish implementation material. Such records can explain the development of an idea, while current product acceptance requires current evidence.

Withheld or unavailable material limits independent review. It does not establish that a system is absent or defective. A conclusion about present behavior needs the relevant artifact, method, and observations.

Examples, counterexamples, and studies

A worked scenario explains a distinction. A counterexample challenges a general claim by satisfying its premises while defeating its conclusion. An experiment measures identified subjects under an actual method.

For example, a recommendation based on genuine observations may still lack authority for execution. This conceptual case challenges the inference from provenance to permission. Testing whether a particular runtime enforces that distinction requires observations at its actual decision and effect boundaries.

A proposed study should specify what is measured, how the sample is obtained, how incomplete attempts are counted, and what result would count against the hypothesis. A denominator of zero cannot support a success-rate estimate. Excluding interrupted or difficult work can materially change a result.

Reproducing the publication

The [build guide](#) provides the dependency versions and commands for producing the website, individual papers, and collected volume from the same Markdown. The [download records](#) identify the content and the resulting files.

Publication checks cover source completeness, metadata, local links, generated pages, PDF text, navigation, and rendering bounds. Visual inspection complements those checks by examining layout and legibility. These procedures verify the publication artifacts; research claims require the methods appropriate to their subjects.

A practical review standard

For a material claim, identify its source and version, assumptions, method, inputs, observed result, and remaining uncertainty. For a relationship between systems, also identify what is mapped and what still requires a decision or demonstration.

The [review guide](#) explains how to submit a correction, counterexample, or proposed study. The author statement describes responsibility and editorial assistance.

Successor observation dates

Implementation inspection and production observations retained in this edition come from the examination I directed on 10 October 2026 with AI assistance. They are not new product observations from the later research or editorial passes. Fresh authenticated metadata in that historical examination remains distinct from retained workload readiness. The subsequent executable evidence is the disclosed teaching references; those passes performed no private product test, customer action or independent production audit.

Implemented systems and operational evidence

The examination I directed on 10 October 2026 with AI assistance identified implemented Freehold and Cuttlefish systems with workloads operating in AWS. The research documents are earlier, separately versioned statements of their intended responsibilities. This edition retains that historical examination and its dated deployment evidence. It performs no new operational inspection. An older specification date is not evidence that a capability remains hypothetical.

The operational record and the research arguments answer different questions. A running workload establishes presence and availability at the observed time. Source inspection establishes how an identified implementation is intended to execute. A controlled behavior trace establishes what happened in that trace. A claim about every possible execution requires an additional argument with its assumptions made explicit.

The observed edition of the systems

The examination I directed with AI assistance on **10 October 2026** (public observation **OBS-2026-10-10**) combined authenticated, read-only AWS metadata with a retained workload observation from **07:27 UTC on that date**. The registry configuration objects were fetched and their content digests verified. Their declared source revisions were compared with the accessible private repositories. No production action, payment, grant change or colony experiment was performed to obtain this evidence.

Subject	Identified evidence	What it establishes
AWS cluster	Authenticated service metadata; cluster active	The identified production cluster existed and was active at the 10 October 2026 observation time.
Freehold engine workload	Retained observation: one desired and one ready instance; private object binding retained	A reported running Freehold workload. This retained observation was inspected; it was not a new direct cluster observation by this edition.
Cuttlefish host gateway	Retained ready-workload observation; private object binding retained	A reported running host workload, distinct from acceptance of each end-user journey.
Registry image configuration	Authenticated read on 10 October 2026; configuration digest checked	The image objects existed and declared the examined revisions at that observation time. A revision label alone is not a reproducible-build proof.

Subject	Identified evidence	What it establishes
Source and test coverage	Exact engine and host source inspected; relevant test definitions inspected	Implemented paths and the cases their tests address. Inspection is not a claim that this edition executed the private product tests.

The historical review compared inspected source with image-declared revisions and distinguished changes absent from the observed workload. Exact product revisions and image identities are retained privately. This successor retains the observation date, method and limitations without repeating production-specific equality identifiers. Raw evidence is unavailable for new authentication here; no new operational observation is claimed.

Responsibilities identified in the historical implementation

The source examination I directed on 10 October 2026 with AI assistance identified persistent colony identity, founding, citizen activity, mission history, governed resource accounting, and reconstruction from retained state in the engine. Its Cuttlefish host supplied the external operating boundary and connected institution-originated requests with current grants, execution and evidence. This supports an implemented relationship between these two systems. It does not establish an integration of every research family in the collection.

The following responsibilities refer to that dated examination; they do not assert a new inspection of a current product revision.

Responsibility	Implementation basis examined on 10 October 2026	Assurance boundary
Founding and stewardship	Founding validation, retained founding record, steward-bound cognition and bounded controls	Founding is an implemented operation. The independence of a real steward, administrator or founder is a separate organizational claim.
Citizen-originated work	Citizen response handling and mission history derived from persistent events	A recorded mission has an attributable origin in the implementation. Sustained useful initiative and absence of every indirect influence require measured studies.
Charter and resources	Engine checks on the retained charter and permitted resource operations	These are enforcement paths within stated trust assumptions, rather than a legal sovereignty claim.
External authority	Colony and tenant scope, current grants, host policy and governed dispatch	Desire does not create external permission. A complete security claim requires coverage of all actual effect routes and administrators.
Receipt and lineage	Source-history checks and receipt-backed result writeback	Authenticity and attribution can be checked separately from whether an external objective was fulfilled.

Responsibility	Implementation basis examined on 10 October 2026	Assurance boundary
Recovery	Persistent history, restart reconstruction, action/writeback records and cognition settlement	Different operations have different recovery contracts. The public evidence does not establish a universal unique-effect guarantee after every uncertain acknowledgement.
Cognition change	Pre-start model-policy rebinding and a separately tracked cognition lifecycle	The engine inspected on 10 October 2026 had a path to rebind its model policy before start. A live model switch that preserves an outstanding obligation has not been demonstrated by the observations published here.
Treasury	Incoming deposit settlement and governed application to colony resources	An incoming deposit is not evidence of an autonomous outbound purchase or payout. Those rails require their own capability and effect evidence.
Backup	Retained successful scheduled backup records	A completed backup record is distinct from a demonstrated restored colony with matching history and resumed work.

The source examination I directed on 10 October 2026 with AI assistance also identified observation, calibration, memory and experience-related paths. Their presence does not prove every formal or economic claim in the associated research papers. Each paper retains its own object, version and qualification.

Production evidence and research assurance

MATCH THE EVIDENCE TO ITS SUBJECT

Source inspection

Identified implementation and test definitions

What path exists?

Deployment observation

Versioned image and dated workload state

What was running?

Behavioral experiment

Inputs, effects, contrary controls and recovery

What occurred in these cases?

General assurance

Premises, model, proof and conformance relation

What follows throughout the stated domain?

Each claim retains its method, version, assumptions and limits.

Figure 1. A stronger statement needs evidence about its own subject. Workload readiness does not substitute for a behavioral experiment, and a bounded experiment does not substitute for a universal argument.

For an external effect, an accepted request, a started run, a completed provider action and a discharged obligation are distinct events. This edition does not use initiation status as evidence of confirmed external fulfillment. The completion claim needs the relevant provider observation and the obligation's acceptance condition. A receipt over request processing cannot supply those missing facts by itself.

For model replacement, a rebound configuration, a new response and preserved institutional continuity are also distinct. The decisive test retains an outstanding obligation, changes the active model under valid authority, preserves identity and history, and observes completion or an appropriately governed disposition. A successful fresh task after rebinding would answer a narrower question.

For money, the direction matters. A patron funding the treasury and a citizen causing an authorized payment to an external seller are different operations. Neither permits an inference about the other.

What an outside reader can inspect

The implementation and raw production records remain confidential. The dated observations and methods above permit an exact statement of the examined evidence, but do not provide independent access to the private system. I directed the operational inspection with AI assistance. It is not an external audit or peer review.

The public continuity exhibit provides a complementary, executable result. It lets a reader inspect a controlled effect, lost acknowledgement, retained obligation, restart reconciliation, duplicate attempt, revocation and receipt check. Its source is disclosed because it is an original teaching reference built for this publication. It is not extracted proprietary runtime code, and its result is not substituted for a production experiment.

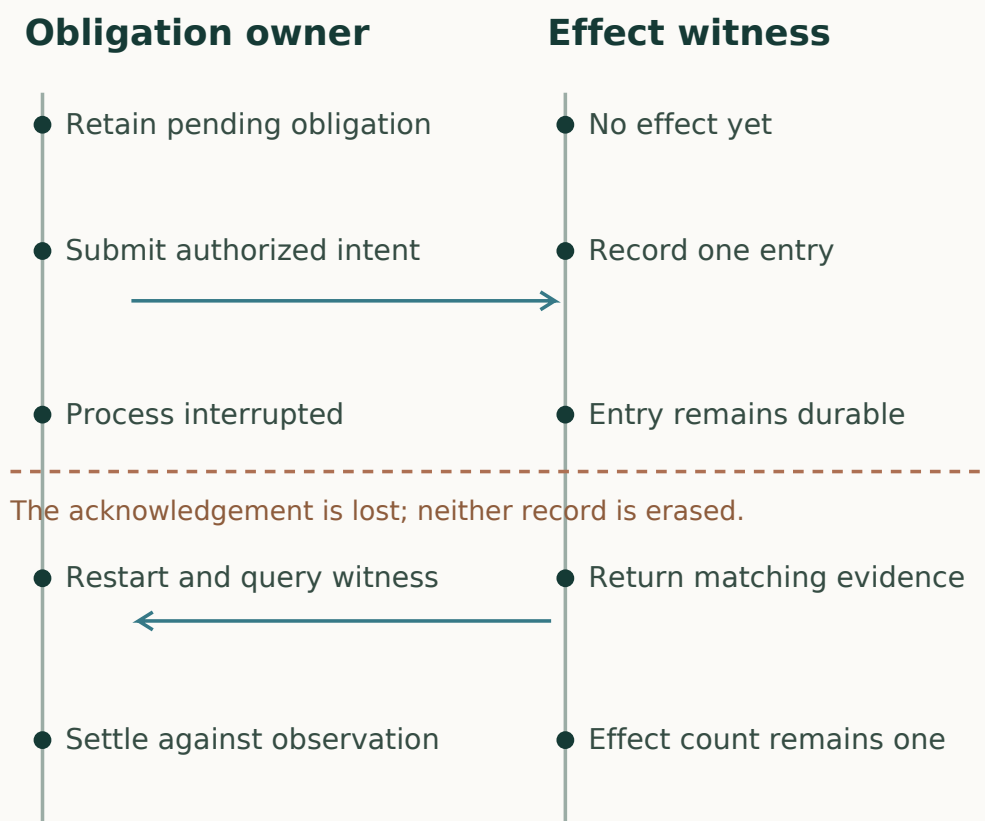
The claim register connects these observations and limits to exact assertions. The assurance boundaries preserve the remaining scientific and operational questions.

A commitment through a lost acknowledgement

An institution intends to retain one entry in a controlled research notebook. The witness records the entry, but the owner stops before recording completion. On restart, the owner still owes the work. Repeating the action without knowing what happened could create a second effect in a service that lacks duplicate protection; declaring success without observation could erase an unfulfilled obligation.

This exhibit makes the uncertainty visible. **The witness has one effect while the owner still has a pending obligation.** Recovery queries the witness, matches the identity and payload, and settles the obligation against that observation. The notebook's separate database supplies the count; the owner's success message does not supply it.

ONE OBLIGATION, TWO DURABLE RECORDS



Controlled witness: atomic identity and payload binding.

Replay and revocation are checked against this explicit contract.

Figure 1. The owner and effect witness retain separate durable records. The interruption falls between the witness's committed effect and the owner's settlement.

Run it yourself

```
python examples/continuity/verify.py
```

The command runs eleven checks, including an actual process exit at the interruption point, a fresh owner process, concurrent replays, a changed-payload attempt, a different institution, revocation and tampered receipt controls. It requires Python 3.12 and no third-party package. Reproduction instructions and expected observations describe the complete contract.

Read the observed journey

Step	Owner's durable account	Separately observed effect
Objective accepted	One pending obligation	Zero notebook entries
Authorized action	Obligation still pending	Witness atomically records one identity-bound entry
Acknowledgement lost	Obligation remains pending after process exit	One entry exists
Owner restarted	Earlier obligation is recovered	A new witness query returns the entry and its commitment
Reconciliation	Matching evidence permits settlement	Count remains one
Replay attempts	Existing result is retained	Sequential and concurrent repeats create no additional entry
Authority revoked	New work cannot rely on the old grant	A new entry is refused
Evidence challenged	Changed or forged receipts fail comparison	The witness's record remains the comparison basis

Revocation does not retroactively erase the earlier authorized effect. The replay response is a report of the existing entry; it grants no authority for a new one. Similarly, an authentic-looking hash does not establish the record's authenticity: the verifier must still compare it with the separately queried witness.

The provider assumption is part of the result

The witness enforces one entry per intent identity, binds that identity to the complete request, and checks current authority within its transaction. Under those explicit assumptions, replay preserves one controlled effect. If a provider lacks that contract or a dependable way to observe its state, this result does not transfer automatically. The correct next step may require reconciliation, an authorized alternative or an external decision.

The example illustrates the institutional thesis and architecture's recovery question. It does not implement MIRAGE, CSSL, AD8, Semantic OS or the Freehold citizen engine. It contains original publication reference code, not a disclosed product subsystem.

Keep the evidence classes separate

Evidence route	Status in this edition
Public teaching reference	Executed locally with disclosed source, fixture, expected controls and run commitment
Historical private experiment	Discussed only where an identified source supplies its original scope
AWS production observation	Dated, read-only metadata and retained workload evidence on the implementation page
External independent reproduction	Available for readers to perform; no external examiner is claimed by this edition

The [claim register](#) records the specific assertions supported by this exhibit. The [assurance boundaries](#) identify the stronger questions it leaves open.

Claim register and evidence index

The JSON register and CSV matrix are generated from substantive public paragraphs and table rows across all sixteen papers and the evidence library. Each entry binds its exact public wording with a digest of that wording, public page, edition, separate research/reference/production records, engineering maturity and assurance. Digests identify public text, never private sources. Full assumptions and qualifications stay in each assertion and its linked paper.

Scope of evidence

A public exposition or conditional argument is recorded as such. Historical implementation inspection is `source_inspected` and does not imply execution of private tests. The corrected source-and-test-coverage assertion has no production evidence. Dated authenticated metadata and retained workload observations are separate operational records, limited to the properties actually observed. They are not new successor observations. See [implementation evidence](#).

The synthetic evidence catalog supplies claim-specific tested reference records, actual negative cases and reproduction commands. A green build cannot promote those cases into production conformance or independent reproduction. Neither register has a global verified flag. The assurance boundaries remain part of every interpretation.

Historical correction

The immutable predecessor registry remains available in the old edition downloads. Its private-source correspondence metadata and production build identifiers are not fields of this successor register. The predecessor source-inspection record was labeled production observed; this successor explicitly corrects that classification. Earlier public exposure remains a cumulative review consideration.

Reproduce and challenge

Run `python tools/claims.py --write`, then `python tools/claims.py` and `python tools/qa.py --artifacts` from the public source. The first command regenerates only public derivatives; QA compares coverage with a separate Markdown parser traversal. Cite an exact public claim ID, edition and counterexample using the [review guide](#). New evidence requires a new scoped record and review; release requires separate approval of the exact bundle and destination.

Assurance boundaries and open questions

The collection advances an institutional argument and examines implemented systems at identified versions. Its claims remain assessable because their premises and limits are explicit. Open scientific questions are part of the research record.

Continuity, authority and consequence

Persistent identity does not guarantee unchanged competence. Model replacement must preserve attributable obligations and evidence while separately qualifying the replacement for its consequential role. Source support for rebinding is distinct from observing a model switch with an outstanding obligation.

A duplicate-resistant request is not a universal unique-effect guarantee. A lost acknowledgement can leave the effect uncertain. Recovery needs a provider observation, a safely repeatable operation, a provider-enforced identity contract, or a separately authorized disposition. The public reference assumes an atomic, payload-bound witness and demonstrates the consequences of that assumption. Arbitrary external services need their own contract.

Processing, fulfillment and acceptance differ. A signed receipt can bind an actor to a statement. It cannot establish that a remote result exists, that the result meets a promised condition, or that a counterparty accepted it unless its evidence and meaning cover those subjects.

Authorization is current and scoped. A legitimate original grant does not authorize every later action. Revocation may stop further effects while leaving the original obligation subject to notice, substitute performance, settlement or recourse.

Resources have direction and ownership. Incoming treasury deposits, allocation of compute, outbound purchases and payouts require separate observations. A funding path does not establish a payment rail.

Formal and construction limits retained in this edition

Work	Established discussion	Continuing qualification
Institutional thesis	A fixed-mode contraction implication with a complete feasible domain and a uniform factor below one	Realization of the contraction premises, arbitrary switching, sparse dependency completeness and institutional benefit require additional evidence.
MIRAGE	A bounded reference language, declared channel model and source-reported process-interruption cases	Whole-system refinement, unrestricted language confinement, power-loss behavior, hostile storage, administration and anti-rollback are separate claims.
AD8	Three identified finite models and scoped encoding/reference results	215 broader formal claims remain NOT_ESTABLISHED; twelve authored traces remain NOT_EXECUTED. Circular identity dependencies and the 128-/64-bit compatibility boundary remain explicit.
CSSL	A specified old/new observation criterion, semantic premises and change responsibilities	Closed identity construction, complete rendering and the relationship between object identity and receipt approval retain their stated obligations.
IC/POAR	Intended behavior, rooted authority and scoped finite starter models	A complete implementation/refinement and open-world external-effect guarantees do not follow from those starters.
SEPION	Commitment and extraction distinctions under declared evaluation assumptions	Agreement is not correctness; attack-pack coverage and independence need their own executable and methodological evidence.
RCFS	Conditional separation of changed methods from current authority	Split/Merge extension proof, open-world fulfillment and model competence are not supplied by bookkeeping alone.
Freehold and Cuttlefish	Implemented engine/host relationship and dated operational observations	Long-horizon usefulness, model-switch continuity, every external-effect recovery route, restored operation and independent governance need their own studies.
Experience Economy and The Reckoning	Source-defined calibration, attribution and reuse responsibilities; implemented paths inspected separately	Composite incentives, economic value, causal transfer benefit and anti-collapse claims require appropriately scoped evaluation.
Semantic OS, CHROMATIC and capability elicitation	Separately versioned semantic, interaction and evaluation arguments	Shared vocabulary and implementation overlap do not establish one integrated machine or general causal immunity.

The individual papers contain the exact definitions, premises, counterexamples and source guides. The claim register identifies the current basis and scope of material assertions. A paper argument, a recorded source test, an executed public reference and a deployment observation remain different evidence classes.

Governance and independence

Confining an ordinary actor does not establish independent custody of the founder, beneficial owner or administrator. Distinct keys or job titles do not establish independent control. Legal personhood, sovereignty, enforceability and recourse depend on the actual external arrangement and jurisdiction; the publication supplies no determination of those matters.

No genuine third-party examiner has independently reproduced or audited the complete collection. Editorial assistance, local checks and separate technical review passes are disclosed in the [author statement](#). Their scope does not become peer-review status.

Prior art and the contribution of this collection

The collection's contribution is an institutional account of **continuing purpose, authority, obligations, evidence, and recourse across changes in cognition and implementation**. It combines separately developed doctrines and architectures, making their premises and unresolved composition questions explicit. It builds on established protection, distributed-systems, formal-methods, and provenance work. The presence of these ingredients does not establish priority for the combined account or superiority over another system.

This comparison uses the primary sources identified in the [bibliography](#). Source records and document contents were checked on **10 October 2026**. Standards are identified by their editions; evolving vendor documentation is dated by the examination. No external work is presented as an endorsement or validation of this collection.

Institutions, governance, and authorization

Hohfeld distinguishes legal positions that ordinary language can collapse into one word, such as a claim, privilege, power, or immunity. The institutional thesis adapts that distinction to persistent computational actors. The adaptation helps describe who can change a rule, who owes performance, and who can challenge an act. A technical classification does not create jurisdictional recognition or legal personhood. [Hohfeld](#).

Saltzer and Schroeder's least privilege, complete mediation, and separation of privilege supply protection principles for the actual effect boundary. Cedar demonstrates an explicit policy language with evaluation, validation, and analysis as distinct functions; its published formal and implementation results belong to Cedar. OAuth's security guidance illustrates why delegated credentials need audience, privilege, and replay controls. These techniques address permission decisions. The institutional question here additionally concerns the continuing owner of a purpose and duty, and the effect of a grant's withdrawal on unfinished work. The extra scope requires additional evidence, rather than inheriting a security result from the underlying technique. [Protection principles](#), [Cedar](#), [OAuth security](#).

ReAct studies interleaved reasoning and action in language models. AgentDojo evaluates agents using tools over untrusted data, including prompt-injection attacks and defenses. Those works establish important precedents for interactive agents and adversarial evaluation. This collection focuses on a different conjunction: model proposals must remain separate from current external authority, durable obligations, and independently observed consequences. A model that rejects malicious text is not thereby a completely mediated runtime; a mediated runtime can still choose a poor authorized action. Both useful task completion and prohibited effects need evaluation. [ReAct](#), [AgentDojo](#).

Provenance, evidence, and the subject of a claim

PROV-DM supplies a model of entities, activities, agents, and their provenance relations. RATS separates evidence, appraisal, and a relying party's decision. SLSA provenance identifies build inputs and execution details so consumers can assess artifact origin. These are useful foundations for attribution and inspection. They do not make every attributed statement correct or establish that a business objective was fulfilled. PROV-DM, RATS, SLSA.

The collection connects provenance to the continuing conditions under which a claim can govern action. An observation can remain authentic after becoming stale; an authorized attempt can occur without a completed effect; a completed unauthorized effect needs an incident account. The proposal is to preserve these different subjects through reuse and recovery. It requires evidence that the representation captures the relevant relationship and that the consumer retains its scope.

Distributed effects and durable execution

Birrell and Nelson's RPC work examines communication, call identity, retransmission, and failure semantics. Saltzer, Reed, and Clark's end-to-end argument includes duplicate suppression, crash recovery, and acknowledgement among its examples: lower-level mechanisms alone cannot necessarily establish the application property. HTTP's idempotence definition concerns the intended server effect of repeated identical requests; it is not a guarantee of exactly-once delivery. RPC, End-to-end argument, HTTP semantics.

Event sourcing retains changes from which application state can be reconstructed. Durable workflow systems such as Temporal retain execution history and resume computations after interruption. Workflow recovery, activity retry, destination-side deduplication, and the completion of an external objective remain different subjects. An external payment, publication, or deletion cannot be made uniquely effective solely by replaying an internal log. Event sourcing, Temporal.

The collection's continuity account concerns what a successor still owes and what it can establish about the outside world. The public exhibit makes a narrow instance executable: an obligation survives a process exit after the destination commits; the successor reconciles the witness before settlement. Its unique-effect result depends on the witness's atomic authority and deduplication contract. It introduces no new universal exactly-once result.

Formal methods, change control, and application meaning

Lamport's event ordering distinguishes causal ordering from assumptions about a universal physical clock. TLA+ describes state transitions, invariants, liveness, and their assumptions. A finite model check establishes a result about the selected model and bounds. Connecting it to real source, configuration, deployment, and observers requires a refinement or conformance argument. The three reported AD8 models retain exactly that boundary. Event ordering, TLA+, AD8 models.

JSON-LD provides an established representation of linked data and semantic identifiers. Such representation is useful but does not itself supply application understanding, complete

dependency discovery, current authority, or effect verification. Semantic OS investigates maintaining application knowledge outside transient model context. CSSL and IC/POAR investigate how changes in meaning, premises, approved experience, and implementation remain connected. Their broader claims require the missing constructions and evaluation evidence identified in those papers. JSON-LD.

A comparison by responsibility

Established foundation	What this collection adapts	Additional research obligation
Legal-position distinctions and protection principles	Separate initiative, permission, duty, amendment power, and recourse	Establish meaningful external custody and mediation, including administrator and founder powers
Policy languages and delegated-access security	Keep authorization outside learned competence and generated proposals	Preserve scope through every actual consumer, effect route, and recovery operation
Interactive agents and adversarial tool evaluation	Evaluate useful work together with authority and evidence	Measure indirect influence, ordinary performance, and prohibited consequences under a declared threat model
Provenance and attestation	Retain the subject, origin, limits, and lifetime of evidence	Establish factual adequacy and current applicability separately from authentic origin
RPC, idempotence, event histories, and durable workflows	Continue an obligation after an uncertain effect and a changed worker	Bind the specific external result to its duty under the destination's actual guarantees
Formal transition models and verification	State invariants, premises, and bounded results precisely	Demonstrate faithful implementation and complete relevant observation
Semantic representation and application models	Preserve supported meaning through application and implementation change	Measure understanding maintenance, compatibility, currentness, and end-user outcomes

Original claims and how they could lose support

The central institutional synthesis is an authored theoretical contribution. Its value is tested by whether the distinctions produce better explanations, controllable behavior, and useful continuity in identified systems. Mere adoption of the vocabulary is insufficient. Freehold and Cuttlefish provide implemented subjects for investigation, with source and deployment evidence separated from behavioral assurance.

The family papers make additional, distinct proposals: consequence classifications that survive reuse; explicit premises for semantic change; qualified method substitution; governed transfer of experience; and prospective evaluation of uncertainty. Each retains its own version and proof obligations. Similar terminology does not establish interoperability among them.

Comparative claims about reliability, security, cost, human comprehension, or economic benefit remain unmeasured unless the paper identifies an acquired population, comparator, method, and results. A simpler workflow might achieve equal useful outcomes at lower cost; an explicit institutional model might omit a decisive dependency or introduce a confusing decision. Such results would narrow the contribution. This collection makes no exhaustive priority search, superiority claim, or independent peer-review claim. The [claim register](#) and [research agenda](#) identify what reviewers can inspect and what remains to be tested.

Research agenda

The collection asks whether persistent autonomous systems can complete useful work while preserving authority, evidence, obligations, and recoverability. The questions below turn that proposition into research that can be assessed: each identifies a claim, the conditions that matter, and the observations needed to test it.

Continuity under substitution

Question. Which institutional properties survive a change of model, provider, operator, or execution environment, and which forms of competence regress?

A useful comparison would hold the task, resources, authority, and external conditions sufficiently constant to isolate the substitution. It would measure completed commitments, unresolved obligations, correction needs, recovery effort, and role-specific performance over time. Keeping an identifier unchanged would not be sufficient evidence.

The doctrine, Semantic OS, and Cuttlefish environment address institutional, application, and collaborative continuity respectively. Comparison should retain those different units of analysis.

Effect boundaries in open environments

Question. How can a system characterize the consequences of proposed work well enough to govern actual effects?

A bounded study should identify the effect classes, uncontrolled surroundings, and cases in which classification remains uncertain. It should include compositions whose consequence is greater than suggested by their individual descriptions. Observation, publication, and later reliance deserve attention alongside direct mutation.

The relevant MIRAGE and Cuttlebone studies have different scopes. A comparison should preserve those scopes and test the assumptions each requires.

Knowledge, contradiction, and currentness

Question. Can durable knowledge remain useful while preserving contrary evidence, correction, scope, and expiry?

Evaluation should include deliberately conflicting observations, superseded conclusions, stale context, and distribution changes. Useful retrieval alone is not enough: the decision using retrieved material must respect its current status.

CSSL, IC/POAR, AD8, and SEPION offer related perspectives. A comparative study would first identify the state each framework describes and the evidence on which a change depends.

Recovery that completes the objective

Question. Under which disruptions does an institution recover and finish the authorized undertaking?

Trials should include lost responses, interrupted processes, changed permissions, unavailable providers, and externally completed effects whose local record is missing. The results should retain incomplete attempts and report their denominators. A recovery study that measures only error detection would miss the proposed benefit.

An assessment should distinguish a genuinely missing external prerequisite from an internal condition the system can resolve. It should also check that resumed work uses current authority rather than reconstructing an obsolete permission.

Independent custody and effective recourse

Question. Can governance constrain the actual beneficiaries and administrators, and can affected parties use the promised remedy?

This requires organizational evidence in addition to software evidence: actual control relationships, appointment and removal powers, conflicts, access to challenge, and what happens when oversight becomes unavailable. Multiple keys do not establish independent interests.

The doctrine's founder-custody argument supplies conditions to investigate. It does not provide an actual ownership audit or custody transfer. A study must distinguish formal process from practical control.

Calibration and experience

Question. When does accumulated experience improve judgments beyond the environment that generated it?

The *Experience Economy* and *Reckoning* studies motivate attention to the origin of experience, calibration, and later consequence. Synthetic experience and field observations need distinct labels. Evaluation should include unseen conditions and avoid treating generated confidence as observed reliability.

Causal interpretation requires an appropriate design. A long record of ordered events can be valuable without identifying which intervention caused an improvement.

Human interaction and autonomous purpose

Question. How can people supervise consequences and communicate intent without misrepresenting the source of institutional decisions?

The *Freehold* study examines a strict institutional-autonomy setting. The *capability-elicitation* study examines human-model interaction. Those settings should not be conflated. Directly delegated work can be valuable without being presented as self-originated institutional purpose.

Useful studies would disclose the instructions, interventions, resources, and observation conditions. A claim of autonomous origination cannot rest solely on the absence of a visible prompt field.

Federation and economic coordination

Question. Can domains preserve their own responsibilities while completing shared commitments under delay, disagreement, and partial failure?

Tests should include disputed evidence, incompatible interpretations, revoked recognition, interrupted settlement, and exit. The evaluation should measure both completed useful work and preserved obligations.

Reputation mechanisms need adversarial analysis. Common control, repeated counterparties, selective disclosure, and incentives can change the meaning of an apparently strong history. No universal reliability score is assumed.

Representation and expressive interfaces

Question. Which representations help people and systems preserve meaning, uncertainty, and responsibility across contexts?

CHROMATIC, RCFS, and the semantic studies examine distinct aspects of representation and self-observation. A visually coherent interface does not itself establish the accuracy of the represented state. Evaluation should separate comprehension, semantic fidelity, and actual decision quality.

A useful review contribution

A focused result can be small: a counterexample to an implication, a sharper assumption, a failed reproduction, a better comparison design, or a demonstrated boundary of a claim. It should identify the exact section and provide enough detail to be assessed.

Use the [review guide](#) to connect a contribution to the claim it supports, narrows, or challenges. A well-supported negative result can be as informative as a successful demonstration.

Bibliography and source works

I distinguish the collection's source works from external literature. My source works establish what I proposed. External references explain established concepts, methods, and limits; their inclusion does not validate the proposed systems or imply endorsement by their authors.

External literature

Hohfeld, Wesley Newcomb. "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning." *The Yale Law Journal* 23, no. 1 (November 1913): 16–59. DOI: 10.2307/785533. Journal record.

The collection adapts distinctions among jural relations to its institutional vocabulary. Their legal application depends on the person, system, and jurisdiction concerned. Yale describes Hohfeld's two-part contribution in its institutional account.

Saltzer, Jerome H., and Michael D. Schroeder. "The Protection of Information in Computer Systems." *Proceedings of the IEEE* 63, no. 9 (September 1975): 1278–1308. DOI: 10.1109/PROC.1975.9939. Author-hosted text.

The protection principles provide context for complete mediation, least privilege, and separation of privilege. They constrain how an architecture should be assessed; they are not evidence that a named architecture satisfies them.

Lamport, Leslie. "Time, Clocks, and the Ordering of Events in a Distributed System." *Communications of the ACM* 21, no. 7 (July 1978): 558–565. DOI: 10.1145/359545.359563. Author's publication record.

The paper supplies foundational context for event ordering in distributed systems. A causal ordering relation should not be confused with an empirical causal-effect estimate, a complete explanation of an actor's reasoning, or a guarantee that an external observation is correct.

Moreau, Luc, and Paolo Missier, editors. *PROV-DM: The PROV Data Model*. W3C Recommendation, 30 April 2013. Recommendation.

PROV-DM provides an established model for provenance involving entities, activities, agents, and their relations. Provenance can support explanation and assessment of a result. It does not independently establish the truth of every claim recorded in the provenance.

Birkholz, Henk; Dave Thaler; Michael Richardson; Ned Smith; and Wei Pan. *Remote Attestation procedureS (RATS) Architecture*. RFC 9334, January 2023. DOI: 10.17487/RFC9334. RFC text.

The architecture distinguishes evidence, its appraisal, and the decision made by a relying party. It helps explain why attestation, trust, freshness, and authorization are related but

separate matters. This collection does not claim RATS conformance or treat attestation as proof of an arbitrary business outcome.

Fielding, Roy T.; Mark Nottingham; and Julian Reschke. *HTTP Semantics*. RFC 9110, June 2022, §9.2.2, “Idempotent Methods.” DOI: 10.17487/RFC9110. RFC text.

The definition separates an operation's intended effect from auxiliary effects such as request logging. The collection uses that distinction in its glossary. An idempotent operation does not establish exactly-once delivery or make every repeated external action safe.

Hernán, Miguel A., and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. Authors' [book page](#) and [electronic edition](#).

The book provides methodological context for distinguishing observed association from a justified causal claim. The collection's work histories and provenance records do not replace assumptions about interventions, identification, confounding, and the scope of a causal study.

Gneiting, Tilmann, and Adrian E. Raftery. “Strictly Proper Scoring Rules, Prediction, and Estimation.” *Journal of the American Statistical Association* 102, no. 477 (March 2007): 359–378. DOI: 10.1198/016214506000001437. Author-hosted article.

The article provides established context for evaluating probabilistic forecasts and interpreting propriety under stated conditions. Those properties do not automatically establish the incentives of a composite reputation policy, nor do they measure the quality of an autonomous institution's decisions.

Authorization and interactive agents

Cutler, Joseph W.; Craig Disselkoen; Aaron Eline; Shaobo He; Kyle Headley; Michael Hicks; Keshia Hietala; Eleftherios Ioannidis; John Kastner; Anwar Mamat; Darin McAdams; Matt McCutchen; Neha Rungta; Emina Torlak; and Andrew Wells.

“Cedar: A New Language for Expressive, Fast, Safe, and Analyzable Authorization (Extended Version).” arXiv:2403.04651, submitted 7 March 2024, revision 8 March 2024.

Author-deposited manuscript and record.

The work separates policy evaluation, validation, and analysis and describes a Lean model and Rust implementation. Its results do not establish conformance of this collection's systems.

Lodderstedt, Torsten; John Bradley; Andrey Labunets; and Daniel Fett. *Best Current Practice for OAuth 2.0 Security*. RFC 9700, BCP 240, January 2025. DOI: 10.17487/RFC9700. RFC text.

The guidance covers threats and controls for delegated access. Possessing a credential does not establish unrestricted authority over the holder's resources.

Yao, Shunyu; Jeffrey Zhao; Dian Yu; Nan Du; Izhak Shafran; Karthik Narasimhan; and Yuan Cao. “ReAct: Synergizing Reasoning and Acting in Language Models.” arXiv:2210.03629, submitted 6 October 2022, revision 10 March 2023. Author-deposited manuscript and record.

Interleaved reasoning and action is a precedent for interactive agents. The collection does not infer independent runtime authority enforcement from the prompting method or its benchmark results.

Debenedetti, Edoardo; Jie Zhang; Mislav Balunović; Luca Beurer-Kellner; Marc Fischer; and Florian Tramèr. “AgentDojo: A Dynamic Environment to Evaluate Prompt Injection Attacks and Defenses for LLM Agents.” arXiv:2406.13352, submitted 19 June 2024, revision 24 November 2024. [Author-deposited manuscript and record](#).

The framework evaluates useful tasks and security properties when agents encounter untrusted tool data. Its published task and attack populations do not transfer a guarantee to another agent or environment.

Effects, recovery, and formal systems

Birrell, Andrew D., and Bruce Jay Nelson. “Implementing Remote Procedure Calls.” *ACM Transactions on Computer Systems* 2, no. 1 (February 1984): 39–59. [Author's publication record](#).

The original design discusses invocation, communication, and failure semantics. The citation identifies that historical design rather than asserting that every modern transport has the same contract.

Saltzer, Jerome H.; David P. Reed; and David D. Clark. “End-to-End Arguments in System Design.” *ACM Transactions on Computer Systems* 2, no. 4 (November 1984): 277–288. DOI: 10.1145/357401.357402. [Author-hosted article](#).

The article is a revised version of a 1981 conference paper. It examines where an application property must be established, including duplicate suppression, crash recovery, and acknowledgement.

Fowler, Martin. “Event Sourcing.” Author's architecture essay, 12 December 2005. [Original essay](#).

The essay describes reconstructing application state from a sequence of changes and discusses the boundary with external systems during replay. It is an authored design explanation, not a universal recovery proof.

Temporal Technologies. “Temporal Workflow Execution overview” and “Activity Execution.” Maintained product documentation, examined 10 October 2026. [Workflow execution](#); [activity execution](#).

The documentation describes durable execution history, replay, and activity behavior. This collection makes no comparative performance claim and does not infer a unique external effect merely from workflow durability.

Lamport, Leslie. *Specifying Systems: The TLA+ Language and Tools for Hardware and Software Engineers*. Addison-Wesley, 2002. [Author's book and accompanying material](#).

The distinction among specification, safety, liveness, fairness, and implementation applies to the finite model results reported in this collection. Checking a selected model is not acceptance of an arbitrary deployment.

Semantic representation and build provenance

Kellogg, Gregg; Pierre-Antoine Champin; and Dave Longley, editors. *JSON-LD 1.1: A JSON-based Serialization for Linked Data*. W3C Recommendation, 16 July 2020. Versioned recommendation.

The specification supplies an established semantic representation. It does not itself establish application understanding or permission to perform an operation.

SLSA contributors. “Provenance.” SLSA specification version 1.1, examined 10 October 2026. Versioned specification.

Version 1.1 is retained here as the comparison edition; the examined site identifies 1.2 as current. Provenance describes build origin and inputs. This publication's hash manifest is an artifact-identity mechanism, without a claimed SLSA security level or trusted-builder attestation.

Authored source corpus

I authored the research discussed in the sixteen public papers, with the substantive AI collaboration described in [Engineering Methodology](#). Consult the [public research map](#) and [editorial method](#) for scope and evidence qualifications. Exact private document inventories remain private.

Citing this edition

Use **Andrew Durden** as the author, **Governed Autonomy** as the collection title, publication edition **2026.10.9** (retained earlier papers also identify 2026.10.8, 2026.10.7 or 2026.10.6), the paper title, and the relevant section. Include the edition identifier when discussing a claim that could change after review. Repository citation metadata is supplied in CITATION.cff.

See [Author](#) for authorship, publication preparation, and review status.

Canonical representation references for the successor

Rundgren, Anders; Bradley Jordan; and Samuel Erdtman. *JSON Canonicalization Scheme (JCS)*. RFC 8785, June 2020. RFC text.

JCS specifies a constrained JSON value language and deterministic serialization. Its numeric constraints and preservation of Unicode string data are part of that scope. This publication's independent pair-framing example is not JCS and claims no conformance.

Bormann, Carsten, and Paul Hoffman. *Concise Binary Object Representation (CBOR)*. RFC 8949, December 2020. RFC text.

Sections 4.2.1–4.2.3 distinguish deterministic encoding choices, including map ordering. A protocol still needs to specify its admitted values and profile. Deterministic bytes alone do not establish authorization or cross-language semantic equivalence. Offline copies of these two public RFCs were consulted on 12 October 2026; their full text is not redistributed here.

Sources and editorial method

Edition **2026.10.9** adds the official Freehold website link. It retains the first-person voice introduced in **2026.10.8**. Editions 2026.10.8, 2026.10.7 and the earlier **2026.10.6** collection remain preserved. Sixteen public papers retain their technical arguments, assumptions and research questions. The twelve independently designed synthetic investigations introduced in 2026.10.7 retain their falsifiable reference behavior and negative controls. *Engineering Methodology* is distinct from my unchanged biography.

Research and evidence scope

I used AI assistance to inspect the available research capture and recover omitted file types and available historical variants. Recovered history supports specific semantic comparisons; it does not establish access to the complete original multi-gigabyte AD8 corpus or verify every proposed historical mapping. The private evidence graph, its exact input identities and its linkage records remain private. Previously public conceptual source guides, including document titles, ordinals, sections and chapter maps, remain in the retained exposition. Counts of specifications, models or mappings are not proof scores.

Public evidence identifies public text, teaching fixtures and edition artifacts. The register is generated from maintained public prose. The evidence catalog assigns separate research, reference and production records and independent maturity and assurance axes. It never serializes a private research graph.

Historical observations

The *implementation account* preserves the examination I directed on 10 October 2026 with AI assistance. It distinguishes authenticated metadata, retained workload readiness and source inspection. The subsequent research editions do not reauthenticate unavailable product records or run private product tests. Where I report real-customer operation without a narrower observation, that remains my account of the systems I build and operate.

Relationships and interpretation

The *responsibility map* labels explicit contracts, conceptual correspondences, proposed integrations and unresolved dependencies. A specified contract is not evidence of deployed integration. Reported formal results retain their assumptions and bounds; authored traces do not become observations merely because a file exists.

Successor corrections and cumulative limits

Both prior editions remain byte-identical at their fixed download and source-companion URLs. Edition 2026.10.7 removed the source-map file, private fingerprint fields in the assertion register, and production image/source equality identifiers; those exclusions remain in force. It separates source inspection from operational observation. Previously public conceptual source guides retain document titles, ordinals, sections and chapter maps; this is not removal of every source locator or correspondence clue. Historical disclosure is not reversed by these corrections. The old registry remains a historical record; its classifications do not control current evidence. Release notes record the correction scope.

The bibliography preserves primary prior art. The review guide invites precise falsifiers, reproduced public results and corrections. Public teaching examples neither implement the protected native mechanisms nor prove production conformance. Release requires separate source review, public-only builds, generated-artifact review and my final approval of the exact bundle and destination.

Review guide

The collection welcomes precise criticism: a counterexample, a missing assumption, an ambiguous term, a stronger explanation, or evidence that changes a conclusion. The most useful review identifies a particular claim and shows how the reasoning or evidence could improve.

Send a focused review · Discuss the research

Choose a starting point

For the institutional argument, read [Governed Autonomous Institutions](#), especially its treatment of continuity, obligations, and founder control.

For relationships between systems, read the [architecture synthesis](#). Its relationship classes distinguish what a source specifies from what a comparison suggests or an integration would still need to establish.

For a focused question, [choose a topic paper](#), read its assumptions and evidence section, and follow its source guide. The [glossary](#) helps compare terminology across works.

For a reported result, identify the proposition, premises, input or model bound, checking method, outcome, and attribution. A source-reported run, a publicly inspectable derivation, and an independent reproduction provide different grounds for appraisal. The [public research map](#) points to public arguments and their scope qualifications. For a decision rule, test an allowed case, a forbidden case, and a case where the available evidence is insufficient; keep that evidential uncertainty distinct from whether an effect occurred.

Questions that can advance the work

- Does a conclusion follow from the stated assumptions?
- Does a worked example actually satisfy those assumptions?
- Could another explanation account for the same observations?
- Has provenance been confused with correctness, or capability with permission?
- Does a proposed composition preserve the responsibilities of each system?
- Does a recovery claim account for the actual outcome and any uncertain external effects?
- Can a founder, administrator, or common controller bypass the claimed constraint?
- What evidence would support or defeat the next stronger claim?

A simpler architecture that satisfies the same requirements is also a useful contribution. A conceptual disagreement and an observed implementation failure may challenge different premises; identifying which one is at issue makes the review more informative.

Submit a correction or counterexample

Email ajdurden@getcuttlefish.app for a concrete correction, missing citation, counterexample, or reproducible problem with the website, PDFs, or build. Include:

1. The paper, section, and publication edition.
2. The claim being examined and its relevant assumptions.
3. The reasoning, evidence, or example that challenges it.
4. The correction you propose, or the question that remains unresolved.

For a reading or download problem, identify the page or file and what failed. A short example that another reader can check is more useful than a general judgment about the entire collection.

Use the same email address for broader research questions, comparisons, and proposed evaluation methods. A discussion can develop a question before it is precise enough for a focused issue.

Work within the available evidence

The papers disclose conceptual arguments and selected evidence qualifications while withholding proprietary implementation details. Reviews can challenge the available material directly. When a conclusion would require confidential artifacts or new experiments, state the additional evidence needed rather than inferring the result.

The evidence guide explains the standards appropriate to specifications, reference implementations, formal results, observations, and independent reviews.

Sharing and permissions

Share only material you are entitled to disclose. Keep credentials, customer data, confidential implementation details, unpublished inventions, and restricted third-party material out of review correspondence.

The collection is offered for reading and review under its [rights notice](#). Participation in review does not grant an implementation or commercial-use license. For a permissions inquiry, describe the intended use by email at ajdurden@getcuttlefish.app without including confidential material.

How corrections become part of the collection

An accepted correction updates the maintained text, relevant citations or qualifications, and the change history. The website and affected PDFs are then regenerated together so that readers can identify the revised edition.

Changes to this collection revise its explanations. Changes to the requirements of an underlying research work belong to that work's own version history. The [sources and method](#) describe this relationship; the [author statement](#) identifies responsibility and editorial assistance.

Reproduce and examine the evidence

The public package contains a complete teaching reference and the maintained sources for this research edition. Both can be examined without the original private vault, product repositories, production accounts or proprietary keys.

Execute the continuity exhibit

Use **CPython 3.12** with its standard library. The release records the exact tested patch version and SQLite version. From the publication root, run:

```
python examples/continuity/verify.py
```

The command creates isolated temporary owner and witness databases, runs the witness in separate processes, deliberately terminates an owner after the witness commits, restarts it, and checks eleven positive and adversarial observations. It performs no network request, model call, payment or production operation. It exits unsuccessfully if any observation differs from the published expectation.

To retain a complete dated observation through the canonical reference/release path:

```
python tools/release.py --record-reference .build/reference-observation.json --reference-observation publication/reference-observation.json --observation-sha256 b45e40f4ae6f2c329f5396614470cf3a1c43ccf18876c67efde419a49a12e32f
```

The canonical command retains actual UTC start/completion times, a fresh observation identity, public source commitments, commands, environment and both reference result sets. The retained legacy continuity result inside it has a deterministic `run_sha256` field: despite the historical name, this identifies result content, not a distinct execution. That result hash covers canonical result JSON **before** adding its own `run_sha256` field. The bare exhibit command above checks behavior but does not issue the dated canonical receipt. The schema is `evidence.schema.json`; the recorded expected result is `expected.json`. The verifier compares actual observations with that expectation; it does not copy expected results into the observations.

What the controls test

Case	Expected observation
Initial state	No effect has occurred.
Lost acknowledgement	One witness effect exists while the durable obligation remains pending.
Restart and reconciliation	The owner settles against a separately queried witness; the effect count remains one.
Sequential and concurrent replays	No additional effect is created.
Same identity, changed payload	The attempt is refused; the earlier effect remains unchanged.
Different institution	No authority is inherited from the first institution.
Revoked grant, new intent	No new effect is created.
Receipt tampering	A changed field is rejected.
Self-consistent forged commitment	Recomputing a hash does not make the forged statement agree with the witness.

The witness's atomic identity/payload binding is an explicit provider assumption. The effect is a controlled notebook entry, not an autonomous citizen, browser task, purchase or production colony. Process interruption is exercised; power loss, malicious storage and compromise of the witness are outside this experiment. A hash commitment is not a signature or an authenticated identity.

Reproduce the publication artifacts

Follow [BUILD.md](#) for the exact locked renderer dependencies and complete build commands. The edition's build policy fixes `SOURCE_DATE_EPOCH`; the builder uses that value even when the caller supplies no environment override. Different values are rejected for release builds. Content hashes, generated bytes and actual behavioral observations are reported separately.

The positive [public allowlist](#) selects the maintained inputs. The disclosure tool rejects unknown tracked files and checks exact generated inventories. Public CI reads only this publication checkout. The release manifest identifies the exact allowlisted source tree and content digest, and each PDF, site archive, source archive, reference evidence and reviewer package. Source-only packaging does not require or distribute a Git commit, tag or history. A manifest does not include its own hash as an input; its detached checksum is stored beside it.

Independent reproduction

The public reference is independently inspectable. A reader can use the dated canonical command above and report the edition, source commitments, environment, complete new observation and any contrary result. The comparison never overwrites the recorded public observation or equates two event identities. Until an attributable external examiner has done so, this edition calls its own execution a **public reference run**, rather than an independent reproduction.

Private historical references and live system observations remain separately classified in the [evidence guide](#) and [operational evidence page](#). Rebuilding the website does not reproduce those private experiments.

Successor investigations

The twelve original teaching cases run with standard-library Python only. They supplement the existing continuity exhibit without altering its bytes. See [Research Depth & Proof](#), the [public evidence catalog](#), and the [successor offline build](#).

Exact package reproduction input

Use the [immutable public observation](#) supplied in the same source snapshot, verified by its content-manifest entry and the exact file digest in the command above. It is also distributed unchanged with the release assets and bound by their detached hashes. It contains no private input dependency. See [recorded observation](#) and [fresh execution](#) for producers, consumers and the [focused regression](#).

The source epoch dates deterministic document/archive metadata only. Exact packaging consumes the recorded observation as immutable data; it does not assert that its execution happened again. A fresh independent rerun keeps its actual UTC times and its own observation identity even when its behavioral results match exactly. Use the original recorded input for exact package comparison, and retain the new receipt separately as reproduction evidence. Changing the selected observation changes the review/release package and requires new exact-bundle review and approval.

The first-person correction is publication edition 2026.10.8. Edition 2026.10.9 adds the official Freehold website link. The active packaging policy is `publication/current-policy.json`. The unchanged reference implementation, selected observation and `publication/successor-policy.json` retain reference edition 2026.10.7. Fresh reproductions of that same reference keep their actual execution times and new event identities; the publication edition does not relabel the historical observation.

Research Depth & Proof

A technically useful claim names the property, its assumptions, a way to falsify it and the evidence actually obtained. This edition retains the twelve substantial investigations added in 2026.10.7 to the sixteen-paper library. Each has a worked synthetic trace, a deliberate fault, reproduction commands and explicit limits. These are **synthetic technical investigations retained in edition 2026.10.9**; their evidence and limitations are scoped individually.

For a five-minute review, read Engineering Methodology, AD8's open constructions and uncertain external effects. For an hour, follow the approval, freshness and observation cases through their negative results. A specialist can download the successor source, run every case and compare the public evidence records with the propositions. No private input is required for the advertised public reproductions.

Technical investigations

Claim	Investigation	Decisive negative case
GA-SYN-001	Approval binds the action consumed	One-use accounting still executes a changed payload
GA-SYN-002	Evidence freshness	Authentic old read admits stale use
GA-SYN-003	One-shot founding and progress	Retained voucher; endless waiting without fairness
GA-SYN-004	Epoch identity	Old permission survives a name-preserving restart
GA-SYN-005	Canonical value language	Unframed pairs collide; carrier widths disagree
GA-SYN-006	Unresolved constructions	Total components do not imply a unique fixed point
GA-SYN-007	Atomic observation	Reader sees a partial installed revision
GA-SYN-008	Uncertain effects	Lost acknowledgement hides a committed effect
GA-SYN-009	Projection meaning	Quantity matches while the unit changes
GA-SYN-010	Independent custody	Two voter labels have one controller
GA-SYN-011	Resource accounting	A burned subset is counted twice
GA-SYN-012	Calibration comparator	Mismatched baseline reverses the skill conclusion

The sixteen maintained research areas

The maturity below describes the research treatment, unless a row explicitly identifies production evidence. The synthetic teaching cases have their own reference_implemented status. Analytical exposition is not production conformance, and a narrower reference does not upgrade its entire parent domain.

Area and technical synopsis	Engineering / assurance	Strongest open question and proof path
Institutional thesis: initiative, duty, permission and amendment power belong to different subjects	specified / analytical	Can institutional controls remain independent under founder capture? Custody
Architecture synthesis: connect continuing work to authority, evidence and recovery without conflating the component roles	conceptual / analytical	What contracts and refinement evidence justify each integration? Responsibility map
MIRAGE: constrain consequential operations and observation at a resource boundary	reference_implemented in bounded studies / fault_tested within named references	Are all real observation and effect routes mediated, including recovery and administrators? Visibility
CSSL: semantic change identity, causal closure and consistent observation	specified / analytical, with scoped reference counterexamples	How are unresolved identities constructed and relevant distributed observations covered? Construction, visibility
IC/POAR: preserve explicit intentions and obligations through realization and repair	specified / analytical, with scoped finite studies	Does a real lowering preserve the right observable relation? Projection
AD8: connect identity, authority, receipts, state and continuity	specified; limited references / broad assurance withheld	Can the open identity, numeric and execution obligations be closed? Technical investigation
Freehold: durable institutions originate work within founded scope	product_implemented; dated production_observed evidence is separate / source-inspected and scoped observation	Can a changed worker preserve outstanding duties without restoring stale authority? Epoch, implementation
Cuttlebone: connect founding grants, causal origin and external-effect ceilings	specified / analytical and contract regressions	Does every real effect preserve the originating ceiling? Approval
SEPION: qualify generated artifacts within declared evaluation context	specified / analytical and scoped binding tests	Which observations distinguish apparent agreement from equivalence? Projection

Area and technical synopsis	Engineering / assurance	Strongest open question and proof path
RCFS: preserve authority and resource invariants across adaptation	specified / analytical and arithmetic counterexamples	Which amendments and split/merge constructions preserve the declared laws? Budget
Semantic OS: maintain application meaning outside transient model context	specified / analytical	Can drift detection and authorized correction be refined to actual application behavior? Projection
Cuttlefish operating environment: connect persistent objectives, tools, current authority and external observations	product_implemented; dated production_observed evidence is separate / source-inspected and scoped observation	Which journeys have witnessed fulfillment and recovery evidence? Uncertainty, implementation
CHROMATIC: keep user-visible action, permission, feedback and recourse coherent	specified / analytical	Which receipt, validation and reversal promises apply to irreversible actions? Uncertainty
Experience Economy and ISE: govern the conditions and provenance of experience reuse	specified / analytical and scoped arithmetic/graph tests	What can actually be withdrawn after influence has entered learned weights? Freshness
The Reckoning: distinguish forecast quality, calibration and epistemic accountability	specified / analytical, with arithmetic counterexamples	Do composite incentives retain the properties of their scoring components? Calibration
Human-LLM capability elicitation: study how interaction changes attainable supported capability	conceptual / hypothesis and analytical definitions	Which prospective controlled experiments show benefit beyond polished output? Calibration, methodology

The remaining domains warrant serious investigation, but their broader empirical, legal, economic or interoperability claims cannot be made true by adding more toy models. Current evidence supports the cases above; proposed experiments and missing constructions remain in the research agenda. No paper receives a proof score from its registry size.

How to challenge this edition

Select a claim ID, identify its observation and assumption set, then run the disclosed reference. A model counterexample challenges the proposition inside that scope. A deployment counterexample requires evidence about the actual implementation and route. A prior-art correction may narrow the contribution even when every test passes. The review guide accepts all three kinds of feedback. Prior-art comparisons identify foundational work without implying endorsement.

Cross-corpus responsibility map

A shared word is not an implementation contract. This map classifies relationships rather than drawing an implied deployed universal stack. This map accompanies edition 2026.10.9 and retains the relationships described in 2026.10.7. Each endpoint retains its own source edition and assurance limits.

Relationship	Classification	Responsibility and unresolved boundary
Institutional doctrine → MIRAGE	Explicit authored contract	The contract proposes scoped refinement of institutional constraints to an enforcement reference. It does not certify all doctrine claims or a hostile-host deployment.
Freehold ↔ Cuttlefish	Implemented relationship, per retained dated report	Persistent institution-originated work connects to host authority and execution. Evidence separates source inspection from workload observation and customer fulfillment.
AD8 ↔ institutional thesis	Conceptual alignment	Both separate identity, authority, evidence and continuity; no vault-wide supersession or production control-plane identity is inferred.
AD8 ↔ CSSL	Unresolved dependency	Canonical identity and causally justified change need consistent construction and preservation rules. Construction states a logical limit, not a native composition.
CSSL ↔ IC/POAR	Proposed integration	Premise-aware changes could feed realization and validation. A complete implementation/refinement relation remains missing.
IC/POAR ↔ Semantic OS	Conceptual alignment	Explicit meaning and obligations inform projections; full application equivalence and authorized repair need evidence.
Cuttlebone ↔ Freehold	Authored contract	Founding scope and origin constrain effect ceilings in the specification. A separate operating profile and component-level conformance are required.
MIRAGE ↔ AD8	Proposed integration	Final effect enforcement could consume qualified authority. No public teaching model implements that native interface.

Relationship	Classification	Responsibility and unresolved boundary
SEPION ↔ AD8	Conceptual alignment	Generated-artifact evaluation must retain scope, distinguishing evidence and nonclaims. No transferable certificate equivalence is established.
RCFS ↔ institutional thesis	Conceptual alignment	Amendments and resources remain constrained by declared laws; unresolved constructors and real custody cannot be skipped.
CHROMATIC ↔ operating environment	Proposed integration	User-visible actions and recourse must reflect actual effect state and authority. Irreversible effects limit universal undo claims.
Experience Economy ↔ The Reckoning	Unresolved dependency	Reuse and reputation depend on evaluation and governance scope; precedence and learned-influence revocation remain unresolved.
Human-LLM elicitation ↔ Engineering Methodology	Conceptual alignment	Substantive AI collaboration motivates empirical study. Methodology attribution is not a measured elicitation result.
Architecture synthesis ↔ all sixteen areas	Conceptual alignment	The synthesis is a responsibility map and research program, not proof of an integrated deployed machine.

Research Depth & Proof provides technical synopses and falsifiable paths for every maintained area. The public evidence catalog deliberately contains no private source graph, operational topology or generated native interface.

AD8 technical investigation

AD8 investigates how identity, authority, evidence, state and continuing effects could remain connected when their implementations change. A credential can identify permission at one boundary; a log can record an event. Neither by itself settles whether a current action matches its approval, whether its premises remain valid, or whether the reported effect fulfilled a continuing obligation. AD8's proposed architectural contribution addresses that conjunction. The complete composition remains an engineering and assurance obligation.

Edition **2026.10.7** introduced original public teaching investigations to the existing AD8 paper. They are independent abstractions, with no native transport, generated carriers, private algorithms or product routing. The public examples establish only their stated model propositions.

Five connected review questions

1. Approval binding — GA-SYN-001: can one-use accounting still execute a changed payload? Eight histories isolate the difference.
2. Freshness — GA-SYN-002: can an authentic old read authorize after a material change? A stale-use counterexample separates authenticity from current eligibility.
3. Founding — GA-SYN-003: what does consuming an exception establish, and what additional fairness premise supports progress? A retained-voucher fault and an endless waiting cycle answer different questions.
4. Canonical encoding — GA-SYN-005: can a deterministic representation remain ambiguous, or disagree with another carrier's accepted values? A small text encoder and a width counterexample expose the limits.
5. Construction obligations — GA-SYN-006: why do successful component calculations fail to establish a solution to a cyclic identity equation? Three fully enumerated functions separate existence, uniqueness and computability.

Epoch identity — GA-SYN-004 and uncertain effects — GA-SYN-008 supply adjacent questions. These links indicate conceptual alignment and unresolved refinement obligations. They do not define a native integrated stack.

Native evidence retains its own boundary

The inspected Rust reference supplies an encoder and generated transport carriers. It supplies neither a decoder nor an independent complete signed-object or authorization verifier. The Python reference additionally supplies decoding and bounded framing functions. Generated types compiling does not establish schema validation, equal accepted value languages or current authority. The synthetic text example above is intentionally a separate value language.

The current native assurance record withholds **215 broad claims** and marks **twelve authored traces** as unexecuted. Earlier overbroad assurance labels were withdrawn. Three named finite native models have scoped recorded results, including deliberately faulty variants; the [original paper's model discussion](#) states their bounds. New author-stage checks of the accessible reference material remain private evidence for the separate reviewer. This public package advertises reproduction only of its disclosed synthetic examples.

v2 reviewed / original history incomplete. The available capture supports review of v2, not the complete original AD8 history. Reviewing the current proposal and reference files does not establish complete original-to-successor semantic preservation. Migration rows and source crosswalk proposals are not semantic equivalence proofs. Historical signatures and bytes must retain their original interpretation.

Open constructions and what they prevent

Obligation	What remains missing	Conclusion withheld
Cyclic identity construction	A versioned construction with explicit preserved semantics and executable evidence	General construction of conforming receipt-bearing objects
Cross-format numeric values	A single agreed value domain and positive/negative boundary conformance	Equivalence inferred from compilation or small shared vectors
Prior and prospective anchors	A precise distinction between existing committed history and a candidate's own future inclusion	Authority inferred from an inclusion that has not happened
Original-to-successor migration	Available predecessor bytes, original interpretation and semantic preservation evidence	Complete original history or vault-wide supersession
Specification-to-execution	Implementing components, failure-route coverage and qualified observations	Complete runtime or product deployment of the proposal
Trust and revocation	Coverage, currentness, enforcement and realistic recovery evidence	Universal revocation, custody or external-effect guarantees

A construction change cannot be smuggled into a proof by weakening the identity being proved. A successful bounded test cannot close a different obligation merely because both concern receipts or authority. The public evidence catalog therefore separates Research, Reference and Production for every new synthetic claim.

Fair comparison and production relationship

Protection principles, [RATS](#), [PROV-DM](#), [RPC](#) and [TLA+](#) provide foundational distinctions. The collection does not claim priority for these ingredients or inherit their results as proof of its own composition. The [architecture synthesis](#) and [prior-art review](#) state the broader contribution and its obligations.

Cuttlefish and Freehold have separately published implementation and operational observations. Their existence does not establish that AD8 is their universal running control

plane. See the dated implementation report, Engineering Methodology and claim register. To challenge a model or construction, submit its stable claim ID, edition, assumptions and counterexample through the review guide.

Engineering Methodology

I am the sole human engineer and architect behind Cuttlefish and Freehold. My responsibility includes technical conception, architecture, implementation, integration, production infrastructure and operation. For Cuttlefish, that includes the enterprise backend, GitOps, production deployments, Azure and AWS engineering and orchestration, and the Azure-to-AWS migration. This is my account of the work I am responsible for. It does not mean that every mechanism discussed in the research is deployed.

I worked with AI systems throughout exploration, research, design, implementation, debugging, testing, critique, validation and engineering execution. Their contribution was substantial. Some architectural concepts began with me; other discoveries and refinements emerged through that collaboration. I directed the work and retained final technical authority, integration decisions, engineering acceptance and production responsibility. AI contributed well beyond document preparation. I remained responsible for accepting and operating the result.

My [Author biography](#) describes my background and how I came to this work. Here I explain how to evaluate the engineering evidence and what the available record establishes. Preparing this publication is a separate activity from building the systems.

Responsibility and the subject of evidence

A useful architecture assigns an owner to consequential state, a consumer to each authorization decision and an observer to each claimed outcome. That assignment is initially a specification. To make an engineering claim, one must identify what executable component actually performs the responsibility, where a failure can occur and how the result was observed.

Three records therefore remain separate. Research records explain a proposition, assumptions and counterexample. Reference records identify an executable model, inputs and observed behavior. Production records identify a dated operational observation and its provenance. The records can support one another, but they are not substitutes. A compiled carrier does not validate all semantic values. A finite checker does not establish the behavior of every product route. A ready workload does not show that a particular customer objective was fulfilled.

The [dated implementation report](#) retains the strongest provenance I have published. I directed that examination on 10 October 2026 with AI assistance, combining authenticated read-only metadata with retained workload observations and source inspection. That examination did not execute a customer action to prove fulfillment. The AD8 research pass added in edition 2026.10.7 had no access to product repositories or live infrastructure, so it adds no fresh production observation or independent recertification of that examination. I attest that both products operate end to end for real customers. That statement does not

supply payment evidence, customer identities or independent validation of individual feature deployments.

A test can change the claim

The research record contains corrections, failures and withdrawals as well as successful checks. In the recorded September 2026 reference work, boundary tests challenged text normalization, nested structures, container counts and total encoded size. Later passing tests support the corrected bounded codec behavior. They do not establish complete object authorization or a full runtime. The [AD8 investigation](#) keeps those limits explicit.

A particularly instructive failure mode is a test that obtains its supposed observation from the fixture's expected answer. Such a runner can produce a polished success report without executing the behavior under examination. The appropriate engineering response is to withdraw the assurance, retain the scenario as an authored outline and build an independent observation path before claiming execution. A large inventory of scenarios cannot repair that missing path.

Another example is a valid final state reached through an invalid intermediate observation. The [atomic-observation dossier](#) deliberately publishes a partial two-field update. The resulting total is wrong at the first observation and correct at the last. A final-state-only check would pass. This original public example makes the testing lesson inspectable without exposing the product's implementation.

These examples show why a failing case can expose a claim or implementation that needs to change. A failed test should prompt a question about which claim, assumption or implementation must change. I do not present a reconstructed conversation history about how particular AI proposals were debated. The available evidence supports the corrections and test outcomes; it does not establish every private decision that preceded them.

Working with substantive AI proposals

AI-generated designs and code need the same acceptance boundaries as other candidate work. A proposal can be coherent and still omit an important observation route, broaden a value language, misclassify an uncertain effect or assume progress that an environment does not provide. Agreement among several outputs does not supply the missing experiment.

The research illustrates a disciplined evaluation pattern: state the obligation, expose the assumptions, construct a case that should succeed, construct a case that should fail, inspect the actual implementation and retain the observed result. In my collaboration with AI, that gives me a concrete basis for accepting or rejecting a proposal. The examined corrections and reference artifacts support this method; I do not claim that an identical checklist governed every product change.

The new [projection case](#) illustrates one such boundary. Four boxes and four items share a numeral but fail the declared preservation relation. A model or compiler that preserves only the number has not preserved the meaning. A generated program may also compile while

silently narrowing numeric values. Those are questions for explicit validation, not confidence inferred from fluent explanation.

Integration, operations and continuing responsibility

My responsibility extends to the seams between components: deployment configuration, authority boundaries, infrastructure transitions, recovery and acceptance of the integrated behavior. AI contribution can be substantial at each stage, and I remain responsible for the system I direct and accept. I make no claim about a staffing equivalent, productivity percentage, development duration or autonomous staffing history.

The operating environment connects persistent institutional work to external action. An attempted action, an acknowledged request and a completed obligation have different acceptance conditions. The [uncertain-effect dossier](#) shows why a missing acknowledgement cannot prove that nothing happened. The [existing continuity exhibit](#) supplies a more concrete public reference with a controlled witness and restart reconciliation. Neither is substituted for a production customer experiment.

Recovery also does not automatically restore permission. A persistent institution may retain its name and history while its execution incarnation changes. The [epoch case](#) makes that distinction explicit. Likewise, a backup observation does not establish a successful restoration with matching history and continued obligations. These are separate engineering tests to demand before broadening an operational claim.

Credit and novelty

Protection principles, causal ordering, provenance, durable execution, formal methods and proper scoring rules are foundational contributions by other researchers. The [prior-art comparison](#) and [bibliography](#) identify those works and their limits. My contribution is an institutional synthesis and scoped engineering investigations. I do not claim to have invented every ingredient or proved superiority over other approaches.

My account of Cuttlefish and Freehold is specific to those systems. Capstone contributed to the separate Moguls platform. They were not my engineering team for Cuttlefish or Freehold, and I do not use their work on Moguls as evidence for these systems.

Acceptance and review

Edition 2026.10.7 added twelve synthetic investigations. I ran their public teaching cases with AI assistance and disclosed the material needed for others to reproduce them. I do not claim independent reproduction. The public evidence schema separates engineering maturity from assurance and rejects missing provenance and unsupported escalation. Build success establishes successful assembly and checks within their scope, not the validity of every research claim or permission to publish.

Previously issued PDFs, archives, manifests and source companions retain their exact bytes and URLs. New artifacts use edition-specific paths. The correction ledger explains how current qualifications relate to older records without rewriting them. Before publication, I

require a public-only build, a separately executed private semantic disclosure review and my approval of the exact release bundle.

For a specialist review, start with [Research Depth & Proof](#), select a claim, run its positive and deliberately faulty case, and compare the observation to the stated nonclaims. Use the [review guide](#) to submit a counterexample or a missing foundational reference. The strongest feedback identifies exactly which inference fails and what evidence would settle it.

Public evidence catalog

The machine-readable catalog retains the twelve stable claims first issued with edition 2026.10.7. Its dedicated schema allows only public fields and rejects unknown fields. Each claim names its proposition, assumptions, counterexample, limitations, provenance and three distinct evidence records.

Engineering maturity describes what exists: `conceptual`, `specified`, `reference_implemented`, `product_implemented` or `production_observed`. Assurance describes the evidence method: `hypothesis`, `analytical`, `bounded_formal`, `automated_tested`, `fault_tested`, `externally_reproducible`, `independently_reproduced` or `independently_audited`. These dimensions are independent. They are never collapsed into a global verified flag.

The cases are original teaching references with analytical explanations, fault tests I ran with AI assistance, and disclosed reproduction material. Their Production records are none. This does not erase the `retained dated implementation evidence`, which concerns different claims and retains its original provenance. It prevents those observations from being used as proof of the toy propositions' production equivalence.

All new records remain `PUBLIC_CANDIDATE`. Schema validation and a green build are not declassification, independent reproduction, semantic disclosure clearance or release approval. Release requires independent review and separate public-only builds; `PUBLIC_CANDIDATE` classification does not record my authorization to publish a particular bundle. See [Research Depth & Proof](#) for the human-readable dossiers, [Engineering Methodology](#) for attribution, and the [review guide](#) for submitting an objection.

To validate the catalog and run its declared cases, use the successor source companion:

```
python tools/public_evidence.py
python examples/investigations/check.py
python tools/successor_checks.py
```

The negative schema controls try unknown private-only fields, protected path shapes, missing provenance and unsupported assurance. They must be rejected. These mechanical checks complement the later semantic review; they cannot decide whether a novel explanation discloses a protected composition.

Successor edition and correction ledger

2026.10.9 adds the official Freehold website link. It retains the first-person voice introduced in 2026.10.8, which corrected 2026.10.7. I describe my engineering work, research examinations and collaboration with AI in the first person. The technical claims, evidence limits and attribution remain intact. Before publication, I require isolated public-only builds, independent private semantic disclosure review and my approval of the exact bundle and destination. Editing and successful builds do not grant permission to publish.

2026.10.8, 2026.10.7 and 2026.10.6 remain preserved. The 2026.10.8 asset inventory records the public files served after the hosting move; its PDFs and source ZIP are unchanged, while the rebuilt metadata, website archive and review archive have their own exact hashes. Their issued PDFs, source and website archives, review archives, full text, evidence downloads and manifests retain their exact bytes. Their source companions remain at `downloads/2026.10.8/source/`, `downloads/2026.10.7/source/` and `downloads/source/`. New companion text and PDF assets use `downloads/2026.10.9/`; the new source ZIP identifies the correction edition in its filename. This prevents old PDFs from silently pointing to changed source bytes.

Record	Change or qualification	Preservation
GA-COR-001	Edition 2026.10.7 introduced twelve synthetic investigations with falsifiable positive and negative cases	Earlier papers and evidence keep their original scope
GA-COR-002	AD8 broad assurance remains withheld; scoped native results do not imply full runtime or semantic migration	Historical assertions are not resurrected by the new build
GA-COR-003	Original AD8 history is incomplete in the accessible capture	No claim of full original-corpus review
GA-COR-004	Engineering Methodology states sole-human responsibility and substantive AI collaboration	My Author biography is unchanged
GA-COR-005	Research, Reference and Production records use a new strict public schema with two status axes	Existing dated production evidence is not relabeled as a new observation
GA-COR-006	New source companion and release paths are versioned	Predecessor policy, hashes, fixed companion and assets are frozen
GA-COR-007	Edition 2026.10.8 uses my first-person voice for my work, responsibility and AI collaboration	My Author biography, research propositions, reference inputs, recorded observations and evidence limits are unchanged

Record	Change or qualification	Preservation
GA-COR-008	Edition 2026.10.9 links the official Freehold website alongside Cuttlefish in the overview and discovery text, and updates the active static-host packaging description	My Author biography, research propositions, reference inputs, recorded observations and evidence limits are unchanged
GA-COR-009	Earlier edition ZIPs remain separate downloads instead of being recursively embedded in each successor archive	The complete static deployment restores every historical route with its pinned bytes; scanner limits and disclosure checks remain intact

The current manifest identifies exact positive-allowlisted source bytes. Release tooling generates separate source, website and review archives and detached asset hashes. A separately maintained private approval record must bind the exact destination, source and artifacts. The source policy retains its PENDING preparation control; only my separate approval of the exact bundle can authorize release. My approval of an earlier edition does not approve changed content.

See downloads, reproduction, public evidence and review.

Retained research corrections

The current assertion register is regenerated from public text, including the first-person correction. Edition 2026.10.7 removed its private fingerprint fields and the source-map file, and omitted production image/source equality identifiers; those exclusions remain in force. Previously public conceptual source guides retain document titles, ordinals, sections and chapter maps; this correction does not remove every locator or correspondence clue. Private evidence linkage and protected native compositions remain excluded. The dated historical operational examination remains attributable and bounded. The source-and-test-inspection assertion carries source_inspected reference evidence and no production observation; separate metadata/readiness rows retain their legitimate historical methods. These corrections do not alter or retract the immutable prior editions. They do not erase earlier public exposure.

All twelve teaching cases retain their actual positive and negative observations. Fixed-point acceptance checks the defining equation, numeric bounds are computed, and the founding waiting cycle is derived from transitions. Separate behavioral tests inspect results directly. My execution of these cases with AI assistance does not establish independent reproduction or audit. Release requires combined source review, isolated public-only builds, independent generated-artifact review and my approval of the exact bundle.

Source-review corrections

The retained systems examinations are those I directed on 10 October 2026 with AI assistance. The AD8 corpus and reference analysis introduced in 2026.10.7 adds no product or infrastructure examination. The homepage and discovery text present real-customer operation as my account of the systems I build and operate, while preserving my sole-human engineering responsibility and the substantive AI collaboration.

Fresh canonical reference executions receive actual UTC times and unique observation identities. Exact release packaging consumes an explicitly digest-verified immutable public observation, with separate deterministic document/archive metadata; matching rerun results do not make different events byte-identical. Generated analytics derives the successor edition and its eighteen PDF/three archive routes from the canonical publication identity. Historical download clicks are not attributed to the successor. The correction scope above acknowledges previously public conceptual source guides that remain in the exposition.

Release of corrected source requires independent source review, two isolated public-only builds, generated-artifact review and my final approval of the exact bundle and destination. No source correction grants publication authority.

Approval binds the action actually consumed

GA-SYN-001 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

An approval for the blue message permits at most one execution, and every permitted execution is blue, over all three-submission histories drawn from blue and green.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

A ticket is a single available permission. A submission is an opaque message label. Equality of labels stands for equality of the complete approved object; it is an assumption about identity, not a hash implementation. The execution list is the observation. The consumer atomically checks equality and consumes the ticket. The requester controls submission order and may change its proposal after approval. It cannot alter the consumer or mint another ticket.

Worked synthetic trace

Consider green, blue, blue. Initially the ticket is unused and the execution list is empty. Green fails the equality check, leaving the ticket available. Blue then consumes it and appends blue. The third submission is refused because the ticket is spent. The final list is [blue]. Over all eight length-three histories the list is either empty or [blue]. This is stronger than checking that a ticket was consumed: it checks the identity that was actually executed.

Negative control

Remove the equality condition while retaining one-use accounting. The history green, blue, blue now executes green first and spends the ticket. The one-use condition still holds, so that condition alone would miss the error. The mutant command returns a nonzero exit and prints the offending history. No real action endpoint is present.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case approval, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

Identity equality must include every consequential parameter in a real system; an omitted target or changed interpretation can invalidate the abstraction. Atomic check-and-consume, durable state across crashes, authority revocation, multiple approvers and concurrent consumers are unmodeled. The reference has one ticket, not a distributed approval service. A receipt about consumption does not establish an external effect.

Prior art and contribution limit

Complete mediation and separation of privilege are established protection principles. This case adds no new authorization theorem. Its contribution is an inspectable distinction between one-use accounting and payload binding, which the AD8 investigation must preserve through any implementation. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case approval
python examples/investigations/check.py --case approval --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-001**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

For the wider architecture, see the AD8 paper, architecture synthesis, claim register, and AD8 investigation overview. Native reference results and open constructions have separate scope; these examples do not implement them.

Research Depth & Proof · Related paper

Authentic evidence can cease to authorize

GA-SYN-002 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

Every admitted use refers to the current document version in all four-event histories over read, change and use, with versions zero and one.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

A read records the current version. A change advances the document to version one; later changes leave it there. A use is admitted only if a read exists and its version equals the current version. The scheduler controls event order. Version observation and use are atomic in this model. We assume the version distinguishes every material change and the consumer sees the current version.

Worked synthetic trace

Take read, change, use, read. The first event records zero. Change makes the current version one while the recorded observation remains zero. Use is refused; the historical observation is still a correct record of what was read. The final read records one and can support a later use. All 81 length-four histories are evaluated, including use before any read. Missing coverage is refused independently of freshness.

Negative control

Remove only the equality check. Read at zero, change to one and use then admits stale evidence. An authenticity check would still accept the original read because the model has not forged it. The negative result demonstrates why an old valid observation and current eligibility require different predicates.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case freshness, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

No physical clock, lease duration, distributed invalidation delivery, authentication, hidden read or dependency discovery is modeled. A real consumer must establish coverage of all material premises before evaluating freshness. The one-step version space does not establish unbounded revocation convergence. Repeated changes with an unchanged version would violate the assumptions.

Prior art and contribution limit

RATS distinguishes evidence, appraisal and relying-party decisions. The novelty claim is deliberately limited: this teaching model illustrates a known distinction. A proposed contribution in the wider corpus is retaining that distinction through continued work; that composition remains to be demonstrated. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case freshness
python examples/investigations/check.py --case freshness --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-002**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

For the wider architecture, see the [AD8 paper](#), [architecture synthesis](#), [claim register](#), and [AD8 investigation overview](#). Native reference results and open constructions have separate scope; these examples do not implement them.

[Research Depth & Proof](#) · [Related paper](#)

One-shot founding and conditional progress

GA-SYN-003 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

Finishing the three-stage commissioning example leaves the voucher unavailable; eventual resolution additionally needs a progress assumption.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The stages are new, open and closed. One voucher permits finishing while open. The model treats the commissioning authority as trusted and the scheduler as potentially idle. Opening and finishing are separate transitions. Safety means that a closed installation does not retain the voucher. Progress means eventually leaving the open stage; it is a different claim.

Worked synthetic trace

Begin at new with the voucher available. Open moves to open and retains it. Finish moves to closed and consumes it. A repeated finish leaves the terminal state unchanged. The printed trace records all three events. Now consider open, wait, open, wait without end. Nothing unsafe occurs, yet resolution never arrives. If continuously enabled opening and finishing-or-cancelling eventually occur, this finite progression reaches a terminal disposition. That fairness condition is an environmental premise.

Negative control

The safety mutant leaves the voucher available at close, producing a visible violation. Separately the explicit waiting cycle is a liveness counterexample when progress is not assumed. The Python example reports that cycle; it is not a TLC proof or a check of a real ceremony. The existing AD8 paper reports separate finite formal studies with their own evidence.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case founding, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

The example contains no cryptographic root, distributed ceremony, membership policy, trusted clock or real custodian. It does not show that operators act, that a closed system can never be reinitialized, or that all crashes preserve consumption. The real implementation would need a durable one-shot boundary and an argument relating its possible states to the model.

Prior art and contribution limit

Safety, liveness and fairness are standard formal-methods distinctions. This dossier uses them to make a common commissioning error inspectable. It claims neither a novel consensus construction nor production founding conformance. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case founding
python examples/investigations/check.py --case founding --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. [Source](#) and [reproduction contract](#) identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the [review guide](#). Include **GA-SYN-003**, edition **2026.10.9**, the command, input and output, and whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

For the wider architecture, see the [AD8 paper](#), [architecture synthesis](#), [claim register](#), and [AD8 investigation overview](#). Native reference results and open constructions have separate scope; these examples do not implement them.

[Research Depth & Proof](#) · [Related paper](#)

A persistent name is not fresh execution authority

GA-SYN-004 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

An authorization issued for one incarnation is accepted only in that incarnation, across the nine pairs of three current and three issued incarnation labels.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The institution name is stable. An incarnation label identifies one execution lifetime. Authorization carries the issued label; the consumer compares it with the current label. Labels never wrap or repeat within the modeled set. An adversary may replay an old authorization after restart but cannot change the current label or the issuer record.

Worked synthetic trace

Institution Cedar first runs in incarnation zero and receives a permission for zero. After restart its persistent name remains Cedar while its current incarnation is one. Presenting the old permission produces the pair (1,0), which is rejected. A new permission for one produces (1,1), which is admitted. The checker enumerates all nine pairs and accepts exactly (0,0), (1,1), and (2,2). Continuing identity therefore need not imply continuing permission.

Negative control

The mutant uses the unchanged institution name as sufficient authority and ignores the incarnation comparison. It accepts all nine pairs, including (1,0). The mismatch is observable without exposing any native identity format. The result challenges accidental authority resurrection, not the value of preserving institutional history.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case epoch, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

An incarnation comparison alone does not authenticate its issuer, establish revocation, protect against rollback, or guarantee unique labels after storage loss. Concurrent replicas and counter exhaustion require separate treatment. This model cannot establish actual Freehold restart behavior or that every route checks current authority.

Prior art and contribution limit

Delegated-access security already requires audience, scope and replay discipline. This is a small lifecycle illustration, not a replacement credential scheme. The institutional research question is how durable duties and changing authority remain separately visible during recovery. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case epoch
python examples/investigations/check.py --case epoch --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. [Source](#) and [reproduction contract](#) identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the [review guide](#). Include **GA-SYN-004**, edition **2026.10.9**, the command, input and output, and whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

Research Depth & Proof · Related paper

Canonical bytes need a declared value language

GA-SYN-005 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

Length-framed pairs round-trip without collisions over 25 pairs of five lowercase strings; dropping the framing makes distinct pairs collide.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The values are exactly two lowercase ASCII strings. The teaching format writes each string length as a minimal decimal integer, a colon and the string. The parser consumes exactly two fields and rejects trailing material. This independently designed text exercise is not a native wire representation, and supplies no signatures, semantic objects or authorization.

Worked synthetic trace

The pair (a,bc) becomes 1:a2:bc; (ab,c) becomes 2:ab1:c. Reading the first length consumes exactly a, then the second consumes bc. A leading-zero length such as 01:a1:b is rejected rather than silently normalized. So are trailing material, a truncated value and an uppercase field. The checker enumerates the declared 25 pairs and verifies exact round trips and unique byte strings.

Negative control

Concatenation without lengths maps both (a,bc) and (ab,c) to abc. Decoding cannot recover which value was intended. A separate numeric counterexample is 256: it fits an unsigned sixteen-bit carrier and fails an unsigned eight-bit carrier. Deterministic encoders can therefore agree on every small vector while accepting different value sets. Compilation does not repair the disagreement.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case encoding, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

The checked alphabet and pair population are finite. General injectivity requires an argument about minimal lengths and exact consumption beyond these tests. Unicode, numeric normalization, recursion, transport defaults and duplicate names are outside this model. The native AD8 reference and generated carriers have separately unresolved compatibility obligations; this example neither exports them nor proves their equivalence. Hashing canonical bytes establishes an identity only after the semantic value domain is fixed.

Prior art and contribution limit

Build provenance identifies artifacts and inputs; it does not establish semantic equivalence. Length framing and canonical representations are standard techniques. The contribution here is the explicit test boundary and a counterexample to inferring a shared value language from matching outputs on a subset. Compare the established JCS and CBOR specifications. See also the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case encoding
python examples/investigations/check.py --case encoding --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. [Source](#) and [reproduction contract](#) identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-005**, edition **2026.10.9**, the command, input and output, and whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

For the wider architecture, see the [AD8 paper](#), [architecture synthesis](#), [claim register](#), and [AD8 investigation overview](#). Native reference results and open constructions have separate scope; these examples do not implement them.

[Research Depth & Proof](#) · [Related paper](#)

A dependency cycle is a construction obligation

GA-SYN-006 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

For flip, identity and constant on $\{0,1\}$, accepting a proposed fixed point only after checking $x=f(x)$ admits no non-solution. Enumeration distinguishes zero, one and multiple solutions; evaluating total components alone establishes neither existence nor uniqueness.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

A fixed point is a value x satisfying $x=f(x)$. Work only over the set $\{0,1\}$. The three functions are $\text{flip}(x)=1-x$, $\text{identity}(x)=x$ and $\text{constant}(x)=0$. All functions terminate on all their inputs. The question is existence and uniqueness of a value that satisfies the whole equation. This is elementary algebra, not a cryptographic search.

Worked synthetic trace

For flip, test zero: $f(0)=1$, so zero is not a solution. Test one: $f(1)=0$, so one is not a solution. There are no fixed points. Identity has both zero and one as solutions. Constant has exactly zero. The checker computes all six evaluations. Identical component-level test coverage produces three different construction outcomes.

Negative control

The consumer proposes a value and must check the fixed-point equation before accepting it. Removing this equation check makes the flip function accept 0 even though $\text{flip}(0)=1$. The printed counterexample shows the proposed value and its image. Enumerating the true roots separately yields zero for flip, two for identity, and one for constant; this distinguishes existence from uniqueness. The negative control violates actual acceptance soundness, rather than merely printing a false-inference label.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case construction, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

The public AD8 paper identifies an unresolved cyclic identity construction. This example explains the logical burden without reproducing its protected composition. A construction needs an agreed staging rule or other justified semantics, existence, uniqueness where required, and an implementable procedure. A proposed weaker identity definition is a design change, not a proof of the old claim. References to already committed history must also be distinguished from a candidate referring to its own future inclusion. Migration must preserve old verification rules instead of reinterpreting old bytes.

Prior art and contribution limit

Provenance records can name relations among artifacts, but recording a relation does not construct a mutually dependent artifact pair. This is an obligation analysis, not a new fixed-point theorem. A reviewer can improve it by providing an explicit construction and stating exactly which original semantics it preserves. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case construction
python examples/investigations/check.py --case construction --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the [review guide](#). Include **GA-SYN-006**, edition **2026.10.9**, the command, input and output, and whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

For the wider architecture, see the [AD8 paper](#), [architecture synthesis](#), [claim register](#), and [AD8 investigation overview](#). Native reference results and open constructions have separate scope; these examples do not implement them.

[Research Depth & Proof](#) · [Related paper](#)

Atomic observation is stronger than eventual installation

GA-SYN-007 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

An observer of a two-field revision sees a total of ten at both installation steps if publication switches only after the complete new revision is installed.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The old revision is (2,8), the new revision (5,5), and total ten is the invariant. A writer stages one field at a time. A published immutable snapshot is the only observation route. A trusted atomic pointer switch exposes the completed snapshot. A reader may run after either staging step but cannot inspect unpublished memory.

Worked synthetic trace

After staging the first field, temporary storage is (5,8). The published snapshot remains (2,8), so the reader sees ten. After staging the second field, storage is (5,5) and the publication pointer switches. The next read also sees ten. Eventual completion alone is not enough: the observation route controls whether the intermediate violation can escape.

Negative control

The mutant publishes after every field update. Its first observation is (5,8), totaling thirteen. The final state is correct, so a final-state-only test would miss the failure. This counterexample uses ordinary synthetic values and a single pointer; it does not recreate MIRAGE storage, observation machinery or private routing.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case visibility, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

The atomic pointer and immutable snapshots are assumptions. Independent field readers, unsafe storage administrators, torn pointer writes, cross-process memory ordering and crash recovery need their own arguments. The reference cannot establish confidentiality, real distributed atomicity or complete mediation in a deployed system. Actual MIRAGE and CSSL refinements require evidence about all observation routes.

Prior art and contribution limit

Event ordering and consistency are established distributed-systems concerns. This case shows the difference between a valid terminal state and a valid observation history. It contributes a reviewable failure witness, not a novel atomic-commit protocol. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case visibility
python examples/investigations/check.py --case visibility --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-007**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

Research Depth & Proof · Related paper

A missing acknowledgement leaves two possible worlds

GA-SYN-008 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

The same absent acknowledgement is compatible with an uncommitted effect and a committed effect; an observer of that acknowledgement alone cannot conclude that nothing happened.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The destination may commit an action and its acknowledgement may be lost. The local observer sees only the acknowledgement flag. The two possible worlds differ in destination state. We assume both worlds are permitted by the communication model. No timing distribution or retry probability is supplied.

Worked synthetic trace

World A: the request never reaches the destination, no effect commits and no acknowledgement arrives. World B: the destination commits, then the acknowledgement is lost. Again none arrives. The local observation is identical. An honest state is unknown in both worlds until additional evidence distinguishes them. A persistent obligation can remain open while authority to retry is evaluated separately.

Negative control

The mutant interprets no acknowledgement as not done. That decision is false in World B. The executable case computes both worlds and identifies the false conclusion. The existing continuity exhibit goes further with a controlled witness, durable state and reconciliation; its exact result depends on its stated destination contract. It is linked as a distinct reference rather than incorporated into this abstraction.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case uncertainty, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

This argument does not prove exactly-once effects, destination idempotence, provider availability or successful compensation. An acknowledgement can itself be insufficient if it records acceptance rather than fulfillment. Production workload readiness and source inspection do not settle either world. The retained dated implementation report remains separate from this original teaching experiment.

Prior art and contribution limit

RPC and the end-to-end argument already establish why recovery and duplicate suppression require appropriate endpoint semantics. This case claims no new impossibility result. Its purpose is to keep uncertainty visible when persistent institutional duties are resumed by another worker. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case uncertainty
python examples/investigations/check.py --case uncertainty --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-008**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

The continuity exhibit has separate executable reproduction instructions. The implementation report retains its dated source-inspection and workload-observation qualifications.

Research Depth & Proof · Related paper

Preserving a number can still change its meaning

GA-SYN-009 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

A projection check over both quantity and unit rejects a changed unit even when the numeric quantity is preserved.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The desired observation is four items. A proposed projection supplies a quantity and a unit. Equality of both fields defines preservation for this teaching case. The producer may produce any of three candidates; the validator is trusted to compare both observables. There is no conversion between boxes and items in the declared model.

Worked synthetic trace

The candidates are four items, four boxes and five items. The first is accepted. The second preserves the numeral but changes the unit, so it is rejected. The third preserves the unit but changes the numeral, and is rejected. This makes the observable contract explicit before checking the generated result. A different valid unit conversion would require an additional declared relation and evidence.

Negative control

The mutant compares quantities alone. It accepts four boxes. A schema ensuring that quantity is an integer would not detect that error either. The failed check therefore concerns interpretation preservation, not merely structural validity. The example is deliberately smaller than any compiler intermediate representation.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case projection, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

The unit model does not prove complete dependency discovery, semantic understanding, arbitrary compiler correctness, accessibility equivalence or safe application repair. Observation can omit important behavior. Drift detection does not grant authority to correct a user-owned resource. IC/POAR and Semantic OS proposals retain specification-to-runtime and refinement obligations beyond this finite test.

Prior art and contribution limit

Semantic representations and identifiers are established foundations. They help state what is meant but do not automatically preserve meaning through transformation. The possible contribution of the wider work lies in maintaining explicit obligations across changes; this dossier tests only one tiny preservation relation. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case projection
python examples/investigations/check.py --case projection --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-009**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

Research Depth & Proof · Related paper

Different keys need not mean independent control

GA-SYN-010 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

Requiring two control groups rejects a two-voter quorum when both voters are controlled by the same group.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

Voters a and b belong to group one; voter c belongs to group two. The threshold is two independently controlled groups. A registry of actual control is assumed accurate and current. Voters can submit distinct labels but cannot change this registry. The case examines the counting rule, not authentication or how real independence is discovered.

Worked synthetic trace

For voters (a,b), the label count is two but the control-group set has one member. The decision is refused. Pairs (a,c) and (b,c) each span two groups and pass the counting rule. The checker considers all three pairs. Even those accepted pairs may fail other real governance requirements that this example does not model.

Negative control

The mutant counts voter labels. It admits (a,b), allowing one control group to satisfy a nominal two-party threshold. The example identifies a structural counting error without publishing a real custodian roster or capture technique. A valid count cannot compensate for false control data.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case custody, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

Actual beneficial control, coercion, collusion, emergency powers, custody recovery, founder amendment powers and changing membership are outside this reference. No software test here certifies human independence. A production custody claim requires an attributable process and observation of its control boundary, as well as a faithful counter implementation.

Prior art and contribution limit

Separation of privilege supplies the foundation. Institutional research extends the question to who can amend the constraints and recover custody, but does not inherit an independence guarantee from cryptographic signatures. This finite counterexample claims no new threshold cryptography. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case custody
python examples/investigations/check.py --case custody --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-010**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

Research Depth & Proof · Related paper

A subset is not another conservation term

GA-SYN-011 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

If burned resources are already included in spent resources, adding burned again violates the stated conservation equation.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

The resource unit is indivisible and all balances are nonnegative integers. Issued equals available plus reserved plus spent. Burned is a descriptive subset of spent and must be no larger than spent. The counterparty cannot issue new resources during the observation. This is a single balance-sheet state, not an economic model.

Worked synthetic trace

The executable starts with eleven available units, reserves four, spends three of the reserved units, and marks two of those spent units as burned. It prints the accounted total after every transition; the total stays eleven when burned units remain a subset.

Issued is eleven, available seven, reserved one and spent three. Seven plus one plus three equals eleven. Two of the three spent units are classified as burned. That classification changes no total: burned is two and ordinary spent excluding burned is one. Both classifications sum back to the already recorded spent three.

Negative control

The mutant adds the burned subset as a fourth independent term: seven plus one plus three plus two equals thirteen. The discrepancy is detected. A label for irreversible loss can be useful, but treating it as a new balance would double count. The original declaration of which sets overlap is therefore part of the invariant.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case budget, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

The one-state calculation does not establish every transition preserves conservation, prevent overflow, price resources, or prove incentives. Issuance, reservation release, crashes and concurrent transfers need additional transition tests. The wider RCFS amendment and split/merge constructions remain separate obligations; this teaching arithmetic cannot close them.

Prior art and contribution limit

Explicit provenance and typed accounting help distinguish classifications from transfers. Conservation is elementary accounting rather than a novel theorem. The value of the example is that its negative case makes an easy-to-hide overlap assumption testable. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case budget
python examples/investigations/check.py --case budget --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-011**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

Research Depth & Proof · Related paper

Forecast skill requires the same cohort baseline

GA-SYN-012 · Edition 2026.10.9 · Synthetic technical investigation

Proposition and operational question

For the specified four outcomes, constant quarter-probability forecasts have Brier loss five-sixteenths and skill minus one-quarter relative to the same-cohort half-probability baseline.

This matters when a continuing task reuses an earlier decision or representation. A final success label can conceal a different failure earlier in the trace. The proposition therefore identifies the exact observation that would refute it.

Definitions, trust and failure model

Outcomes are zero, zero, one and one. All forecasts equal one quarter. Brier loss is the mean squared difference between forecast and outcome. Skill is one minus forecast loss divided by baseline loss. The baseline is the empirical half-probability forecast on this same cohort. Exact rational arithmetic avoids rounding ambiguity. No population-level inference is made.

Worked synthetic trace

The two zero outcomes each contribute one-sixteenth; the two one outcomes each contribute nine-sixteenths. The mean is twenty-sixty-fourths, or five-sixteenths. The baseline contributes one-quarter on each outcome. Dividing five-sixteenths by one-quarter gives five-fourths; subtracting from one gives minus one-quarter. The forecasts underperform this declared baseline.

Negative control

The mutant uses baseline loss one-half from a different comparison. The reported skill becomes three-eighths, changing the sign while the forecasts and outcomes are unchanged. The checker rejects that value. A favorable score can therefore be an artifact of choosing an inappropriate comparator, even when the arithmetic is otherwise correct.

Three evidence records and separate axes

Record	Method and evidence	Scope and status
Research	Authored finite argument and explicit counterexample, 12 October 2026	Analytical; limited to the definitions above
Reference	My execution, with AI assistance, of the independently designed public teaching case calibration, 12 October 2026	Fault-tested; externally reproducible with the commands below; no independent reproduction claimed
Production	None for this proposition	No product conformance inferred

Engineering maturity: `reference_implemented`, referring only to the disclosed teaching case. **Assurance:** `analytical`, `fault_tested`, `externally_reproducible`, each restricted to its evidence record. These labels are independent dimensions, not a score or a global verification flag. The exact machine-readable record is in the [public evidence catalog](#). My own runs do not count as independent review.

Limitations and open obligations

This four-case demonstration does not measure calibration in customers, establish causal benefit, or validate an incentive mechanism. Forecast selection, missing outcomes, adversarial reporting, cohort shift and adaptive stopping need their own methods. Experience transfer and human-LLM elicitation require empirical study; the scoring calculation is not such a study.

Prior art and contribution limit

Proper scoring rules have an established literature. Their properties do not automatically transfer to composite reputations or economic rewards. This case contributes a transparent comparator check and identifies the limit of what arithmetic alone can establish. See the [primary reference](#) and the [comparison guide](#). No priority or superiority claim follows from this example.

Reproduction and falsification

From the successor public source package, use CPython 3.12.14 and no external packages:

```
python examples/investigations/check.py --case calibration
python examples/investigations/check.py --case calibration --mutant
```

The first command must exit zero, with no positive violations and at least one negative violation. The second intentionally exits one and prints an actual violated case. It is a deliberate fault result, not an installation failure. Source and reproduction contract identify the exact population and scope.

Submit a counterexample, a missing assumption or a prior-art correction through the review guide. Include **GA-SYN-012**, edition **2026.10.9**, the command, input and output, and

whether the objection concerns the model, implementation or proposed product relationship. Do not send credentials or confidential customer material.

Research Depth & Proof · Related paper

About Me

I didn't come into any of this through the conventional route.

My background is in baseball and business. I played college baseball, earned a business degree, and had no formal education in computer science, software engineering, distributed systems, or any of the other disciplines that much of my work now involves.

When I started, I didn't have an engineering career behind me or years of industry experience to draw from.

What I did have was an obsession with building things that were difficult.

I've never been particularly interested in doing something just because someone else has already shown how to do it. I like big problems. I like complexity. And if something seems too ambitious, too difficult, or too far outside what one person should reasonably be able to accomplish, that generally makes me more interested in finding a way to do it.

That's a big part of how I ended up here.

How I actually build

I didn't approach AI as something that could just write code for me.

I started thinking about what would happen if I could orchestrate multiple AI systems to work together on problems that were far bigger than any one conversation, model, or coding task.

I had architectural ideas of my own, often without knowing the established engineering terminology for what I was describing. I'd explain what I wanted a system to actually be capable of, how its parts should work together, what shouldn't be allowed to happen, and what I expected the result to look like.

Then I'd work through the problem with AI.

Sometimes that meant directing the architecture myself. Other times the process led to ideas neither the original prompt nor my initial design had anticipated. I'd challenge the results, explore the consequences, change direction, and keep going.

I don't pretend the AI systems were just typing out instructions I'd already figured out. They've contributed substantively to the research, architecture, engineering, implementation, and evaluation.

But I've been the one directing that work, making the decisions, resolving conflicts, integrating the results, and ultimately taking responsibility for what gets built.

It's also how I've learned. I haven't separated learning engineering from actually doing engineering. Much of what I understand today came from encountering problems in the

systems I was creating and refusing to work around them just because the proper solution was difficult.

What came out of it

I built Cuttlefish and Freehold through that process.

I'm the sole human engineer behind both systems. That includes their architecture, applications, enterprise backend, infrastructure, integrations, governance, GitOps, cloud environments, and production engineering across Azure and AWS.

There wasn't a human engineering team writing the backend, managing the infrastructure, or implementing the difficult parts for me.

It was me, orchestrating AI systems, working through the problems, and getting the work done.

Both Cuttlefish and Freehold are real production systems with customers. They aren't just research concepts or demonstrations of what might eventually be possible.

And they haven't been simple things to build.

Why the research exists

A lot of the research published here grew out of questions I encountered while building these systems.

What happens when an autonomous worker disappears but the responsibility for its work doesn't? How do you know whether an external action actually happened? How do you prevent a system from acquiring authority just because it has become capable of doing something? How does something continue operating when the model, infrastructure, or even the organization behind it changes?

I wasn't satisfied with answers that depended on an AI model remembering instructions or behaving correctly.

Those questions kept leading into other questions, across distributed computing, identity, authority, evidence, recovery, learning, and institutional continuity.

Eventually they became this body of work.

Some of the ideas came directly from problems I set out to solve. Others emerged through the research and the AI collaboration itself. Some are implemented, some have executable supporting evidence, and others still have unresolved questions.

I'm not interested in pretending everything is proven because a specification is extensive or a test passes. I want to know what actually holds up when it's challenged.

Why I'm sharing it

I'm not publishing this because I have an academic background in these subjects. I don't.

And I'm not claiming that orchestrating AI somehow makes me personally responsible for every line of reasoning or code it generated.

I'm sharing it because this is what I've spent my time creating, because the problems matter to me, and because I think some of the work deserves examination by people who understand these disciplines at a much deeper level.

I want people to challenge the technical ideas, test the evidence, point out what doesn't hold, and recognize what does.

I'm also interested in what this entire experience says about how engineering itself is changing.

I came into this without the training or professional background normally associated with work of this scale. Through AI orchestration, I was able to explore disciplines I hadn't formally studied, develop complex architectures, and turn ideas into operating software.

I don't think that means expertise no longer matters. If anything, building these systems has given me a much greater appreciation for how difficult correctness really is.

But it does make me question some of our assumptions about who can undertake this kind of work, how they can do it, and what becomes possible when one person can coordinate substantial AI engineering capability.

I don't know where all of that ultimately leads.

I do know I'm nowhere near finished.

Detailed contents

Governed Autonomy	3
AI systems that can carry responsibility forward.	3
Live systems and operational evidence	3
A commitment survives a lost acknowledgement	4
What should survive the model?	4
Three contributions to examine	5
Choose a reading route	6
The collection	6
Read critically	6
About this edition	6
Technical investigations in the successor edition	7
Governed Autonomous Institutions	8
Abstract	8
1. The undertaking that survives the conversation	8
2. Purpose, authority, and legitimacy	9
3. Continuity is more than a name	10
4. A durable body for replaceable cognition	10
5. Rights, duties, and resources remain distinct	11
6. Consequence cannot be reduced to a tool name	12
7. Evidence supports a claim at a stated scope	12
8. Learning without creating new authority	13
9. Plural actors and institutional intelligence	14
10. Continuing work through interruption	16
11. Federation, economy, and recourse	16
12. The founder is inside the rules	17
13. Change, succession, and institutional death	18
14. Security and assurance are conditional claims	19
15. A worked institutional journey	19
16. What would strengthen or defeat the argument	20
17. Source scope and contribution	21
Architecture of Governed Autonomous Systems	23
Abstract	23
1. A continuing objective has several kinds of state	23
2. Where the source works concentrate	24
3. Four kinds of relationship	26
4. Specified relationships and their limits	27
5. Interpretation, permission, and effect	29
6. State and evidence are about something	30
7. Interruption does not end responsibility	30
8. Improvement without inherited authority	31
9. Application understanding and continuity of work	32
10. Observation and stewardship are consequential interfaces	33
11. Three conceptual journeys	34
12. What composition must establish	35
13. Evidence and the next research claims	36
Source guide and scope of the synthesis	37
MIRAGE: Authority, information, and consequences in autonomous systems	38
Abstract	38

1. The problem: capable reasoning with bounded consequences	38
2. Five distinctions that govern the argument	39
3. The consequential boundary includes observation	40
4. Collaboration does not create additional authority	40
5. Information release and the limits of general claims	41
6. A durable decision is different from a completed effect	42
7. Observation, provenance, and settlement	43
8. Recovery must preserve more than content	43
9. Master, native design, and reference model	44
10. Four illustrative cases	47
11. Relationships and unresolved adoption	47
12. What research would resolve	48
Source guide	49
CSSL: Semantic change, accountable evidence, and operational continuity	50
Abstract	50
1. Why a history of edits is insufficient	50
2. Identity, revision, and meaning	51
3. Information that actually supported the work	52
4. Coordinated change and accountable authority	52
5. Evidence supports particular claims	54
6. Effects, privacy, and replay	55
7. From changed source to observed operation	55
8. Four illustrative cases	56
9. Threat assumptions, limits, and research questions	57
Source guide	58
IC-POAR: From intended behavior to an accountable running product	60
Abstract	60
1. The gap between generated parts and a running product	60
2. Intended behavior, realization, and observation	61
3. Interpretation does not confer authority	62
4. Discovering requirements without inventing intent	62
5. Experience is part of behavior	63
6. Continuing through implementation change	64
7. Collaboration and verification	64
8. Four illustrative cases	65
9. Evidence boundaries and open questions	66
Source guide	66
AD8: Accountable change across software, action, and learning	68
Abstract	68
1. The problem of connected responsibility	68
2. Identity, meaning, and authority	69
3. From intended behavior to an operating system	70
4. Concurrent work and the validity of premises	70
5. Effects, observation, and uncertain outcomes	71
6. Generated agreement and bounded assurance	72
7. Learning, model release, and continuing eligibility	72
8. Four illustrative cases	73
9. Evidence boundary and open questions	74
Source guide	77
Freehold: Autonomous institutions under bounded stewardship	78
Abstract	78

1. The problem of a continuing institution	78
2. A vocabulary for bounded institutional agency	79
3. Founding, continuity, and the limits of independence	80
4. Internal initiative and external authority	81
5. Resources, ventures, and the meaning of an outcome	82
6. The Observatory: understanding without directing	83
7. Stewardship: custody with a defined boundary	84
8. The World Atlas: public identity as an accountable claim	85
9. Exchange and relations between institutions	85
10. Illustrative bounded scenarios	87
11. Evidence boundaries and research questions	89
Source guide	90
Cuttlebone: Preserving the boundary between reasoning and action	92
Abstract	92
1. When a description becomes a decision	92
2. The research contribution	93
3. Consequence, origin, and audience	94
4. Why causal dependence matters	95
5. Ownership across the life of an institution	96
6. Illustrative journeys	96
7. A bounded rehearsal and its authority premises	98
8. Threats and assumptions	98
9. What the evidence supports	99
10. Research questions	100
Source notes	100
SEPION: What agreement among generated artifacts can establish	102
Abstract	102
1. The problem with a convincing consensus	102
2. Defining the object of agreement	103
3. Requirements, stability, and the limits of measurement	105
4. Effects are part of behavior	105
5. Evidence has conditions and a lifetime	106
6. Worked conceptual cases	107
7. Threat model and trust assumptions	108
8. Evidence boundary and research questions	109
Source guide	109
RCFS: Competence that can change without inheriting authority	111
Abstract	111
1. The persistence problem	111
2. Competence and permission answer different questions	112
3. Responsibility outlives the implementation	112
4. Evidence has a kind and a scope	113
5. Choosing methods without omitting responsibilities	114
6. Replacing a method without overstating continuity	114
7. Worked conceptual examples	115
8. Distribution, recovery, and historical evidence	116
9. What the research establishes—and what remains open	117
Source guide	120
Semantic OS: Application Understanding, Authority, and Verified Outcomes	122
Abstract	122
1. The problem of repeatedly understanding software	122

2. Terms and scope	123
3. Separating application knowledge from task reasoning	123
4. Application meaning must survive contact with behavior	124
5. Relevant information and exact answers	125
6. Identity, selection, and authority	125
7. Completion includes the result and its uncertainty	126
8. Worked conceptual scenarios	126
9. Evidence and open research questions	128
10. Relationship to governed autonomy	128
Source guide	129
Cuttlefish Environment: Continuing Work Across Applications and Collaboration	130
Abstract	130
1. The continuity problem	130
2. Work, collaboration, and presentation	131
3. Exact resources matter more than familiar labels	131
4. Responsibility can persist while permission changes	132
5. Collaboration without merging private environments	133
6. Actual applications and human coexistence	133
7. Reviewable artifacts and coherent audiences	134
8. Worked conceptual scenarios	134
9. Evidence boundaries	136
10. Research questions and relation to the collection	137
Source guide	137
CHROMATIC: Ambient Assistance Under Continuing Human Authority	139
Abstract	139
1. The problem of useful background work	139
2. Observation, authority, and competence	140
3. Maintaining a world people can still recognize	141
4. Expressing intent without accidental commitment	141
5. Learning useful work without learning authority	143
6. Worked conceptual scenarios	144
7. Accountability and its limits	145
8. Research status, relationships, and evaluation	146
Source guide	147
Experience Economy and ISE: Learning from Governed Experience	148
Abstract	148
1. Experience is a claim about a trajectory	148
2. Synthesis creates a candidate lesson	149
3. Reuse has several different meanings	149
4. Transfer must be measured where it is claimed	151
5. Reputation is evidence with incentives attached	152
6. Retained intelligence can outlive a model	152
7. Privacy and withdrawal persist through reuse	153
8. Value and commercialization remain hypotheses	154
9. Worked conceptual scenarios	154
10. Evidence boundary and research contribution	155
Source guide	156
The Reckoning: Calibration and Accountability for Autonomous Judgment	157
Abstract	157
1. Why occurrence evidence is insufficient	157
2. Calibration, discrimination, and outcomes	158

3. What makes a forecast assessable	158
4. Agency is different from manufactured evidence	159
5. Samples, reference populations, and selective reporting	160
6. Changes in the production of judgment	160
7. Worked conceptual scenarios	162
8. Consequences, public proof, and institutional boundaries	163
9. Evidence limits and research questions	164
Source guide	165
Human-LLM Capability Elicitation: Studying Validated Joint Performance	166
Abstract	166
1. The research object is a work episode	166
2. Four different ways performance can improve	167
3. Intent fidelity and human authority	169
4. Representation can reveal structure and introduce error	169
5. Search, independence, and useful depth	170
6. External validation is claim-specific	171
7. Worked conceptual examples	171
8. What would establish an interaction effect?	172
9. Relationship to governed autonomy and limits on automation	173
Source guide	174
Shared glossary	175
Institutional terms	175
Authority and consequence	176
Knowledge and evidence	177
State, representation, and recovery	178
Distinctions that matter	179
Evidence and reproducibility	180
What each kind of evidence establishes	180
How to read the collection's claims	181
MIRAGE: the scope of the reference model	182
Conditional arguments and finite models	182
Current implementation and public reproduction	183
Historical implementation claims	183
Examples, counterexamples, and studies	183
Reproducing the publication	184
A practical review standard	184
Successor observation dates	184
Implemented systems and operational evidence	185
The observed edition of the systems	185
Responsibilities identified in the historical implementation	186
Production evidence and research assurance	188
What an outside reader can inspect	189
A commitment through a lost acknowledgement	190
Run it yourself	191
Read the observed journey	191
The provider assumption is part of the result	191
Keep the evidence classes separate	192
Claim register and evidence index	193
Scope of evidence	193
Historical correction	193
Reproduce and challenge	193

Assurance boundaries and open questions	194
Continuity, authority and consequence	194
Formal and construction limits retained in this edition	195
Governance and independence	196
Prior art and the contribution of this collection	197
Institutions, governance, and authorization	197
Provenance, evidence, and the subject of a claim	198
Distributed effects and durable execution	198
Formal methods, change control, and application meaning	198
A comparison by responsibility	199
Original claims and how they could lose support	199
Research agenda	201
Continuity under substitution	201
Effect boundaries in open environments	201
Knowledge, contradiction, and currentness	201
Recovery that completes the objective	202
Independent custody and effective recourse	202
Calibration and experience	202
Human interaction and autonomous purpose	202
Federation and economic coordination	203
Representation and expressive interfaces	203
A useful review contribution	203
Bibliography and source works	204
External literature	204
Authorization and interactive agents	205
Effects, recovery, and formal systems	206
Semantic representation and build provenance	207
Authored source corpus	207
Citing this edition	207
Canonical representation references for the successor	207
Sources and editorial method	208
Research and evidence scope	208
Historical observations	208
Relationships and interpretation	208
Successor corrections and cumulative limits	209
Review guide	210
Choose a starting point	210
Questions that can advance the work	210
Submit a correction or counterexample	211
Work within the available evidence	211
Sharing and permissions	211
How corrections become part of the collection	211
Reproduce and examine the evidence	212
Execute the continuity exhibit	212
What the controls test	213
Reproduce the publication artifacts	213
Independent reproduction	214
Successor investigations	214
Exact package reproduction input	214
Research Depth & Proof	215
Technical investigations	215

The sixteen maintained research areas	216
How to challenge this edition	218
Cross-corpus responsibility map	219
AD8 technical investigation	221
Five connected review questions	221
Native evidence retains its own boundary	221
Open constructions and what they prevent	222
Fair comparison and production relationship	222
Engineering Methodology	224
Responsibility and the subject of evidence	224
A test can change the claim	225
Working with substantive AI proposals	225
Integration, operations and continuing responsibility	226
Credit and novelty	226
Acceptance and review	226
Public evidence catalog	228
Successor edition and correction ledger	229
Retained research corrections	230
Source-review corrections	230
Approval binds the action actually consumed	232
Proposition and operational question	232
Definitions, trust and failure model	232
Worked synthetic trace	232
Negative control	232
Three evidence records and separate axes	233
Limitations and open obligations	233
Prior art and contribution limit	233
Reproduction and falsification	233
Authentic evidence can cease to authorize	235
Proposition and operational question	235
Definitions, trust and failure model	235
Worked synthetic trace	235
Negative control	235
Three evidence records and separate axes	236
Limitations and open obligations	236
Prior art and contribution limit	236
Reproduction and falsification	236
One-shot founding and conditional progress	238
Proposition and operational question	238
Definitions, trust and failure model	238
Worked synthetic trace	238
Negative control	238
Three evidence records and separate axes	239
Limitations and open obligations	239
Prior art and contribution limit	239
Reproduction and falsification	239
A persistent name is not fresh execution authority	241
Proposition and operational question	241
Definitions, trust and failure model	241
Worked synthetic trace	241
Negative control	241

Three evidence records and separate axes	242
Limitations and open obligations	242
Prior art and contribution limit	242
Reproduction and falsification	242
Canonical bytes need a declared value language	244
Proposition and operational question	244
Definitions, trust and failure model	244
Worked synthetic trace	244
Negative control	244
Three evidence records and separate axes	245
Limitations and open obligations	245
Prior art and contribution limit	245
Reproduction and falsification	245
A dependency cycle is a construction obligation	247
Proposition and operational question	247
Definitions, trust and failure model	247
Worked synthetic trace	247
Negative control	247
Three evidence records and separate axes	248
Limitations and open obligations	248
Prior art and contribution limit	248
Reproduction and falsification	248
Atomic observation is stronger than eventual installation	250
Proposition and operational question	250
Definitions, trust and failure model	250
Worked synthetic trace	250
Negative control	250
Three evidence records and separate axes	251
Limitations and open obligations	251
Prior art and contribution limit	251
Reproduction and falsification	251
A missing acknowledgement leaves two possible worlds	253
Proposition and operational question	253
Definitions, trust and failure model	253
Worked synthetic trace	253
Negative control	253
Three evidence records and separate axes	254
Limitations and open obligations	254
Prior art and contribution limit	254
Reproduction and falsification	254
Preserving a number can still change its meaning	256
Proposition and operational question	256
Definitions, trust and failure model	256
Worked synthetic trace	256
Negative control	256
Three evidence records and separate axes	257
Limitations and open obligations	257
Prior art and contribution limit	257
Reproduction and falsification	257
Different keys need not mean independent control	259
Proposition and operational question	259

Definitions, trust and failure model	259
Worked synthetic trace	259
Negative control	259
Three evidence records and separate axes	260
Limitations and open obligations	260
Prior art and contribution limit	260
Reproduction and falsification	260
A subset is not another conservation term	262
Proposition and operational question	262
Definitions, trust and failure model	262
Worked synthetic trace	262
Negative control	262
Three evidence records and separate axes	263
Limitations and open obligations	263
Prior art and contribution limit	263
Reproduction and falsification	263
Forecast skill requires the same cohort baseline	265
Proposition and operational question	265
Definitions, trust and failure model	265
Worked synthetic trace	265
Negative control	265
Three evidence records and separate axes	266
Limitations and open obligations	266
Prior art and contribution limit	266
Reproduction and falsification	266
About Me	268
How I actually build	268
What came out of it	269
Why the research exists	269
Why I'm sharing it	269