GOVERNED AUTONOMY

Governed Autonomous Institutions

A paper in the Governed Autonomy collection

Andrew Durden
Independent research
Publication edition 2026.10.9

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

Content identifier: 7150dfae1470e29ab5d0fbcc

All rights reserved · Research exposition

Contents

Governed Autonomous Institutions **3**

Abstract . . . 3

1. The undertaking that survives the conversation . . . 3

2. Purpose, authority, and legitimacy . . . 4

3. Continuity is more than a name . . . 5

4. A durable body for replaceable cognition . . . 5

5. Rights, duties, and resources remain distinct . . . 6

6. Consequence cannot be reduced to a tool name . . . 7

7. Evidence supports a claim at a stated scope . . . 7

8. Learning without creating new authority . . . 8

9. Plural actors and institutional intelligence . . . 9

10. Continuing work through interruption . . . 11

11. Federation, economy, and recourse . . . 11

12. The founder is inside the rules . . . 12

13. Change, succession, and institutional death . . . 13

14. Security and assurance are conditional claims . . . 14

15. A worked institutional journey . . . 14

16. What would strengthen or defeat the argument . . . 15

17. Source scope and contribution . . . 16

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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 $0 \le q < 1$ such that $d(F(x),F(y)) \le q\, d(x,y)$ for every pair of feasible activations. Successive updates then satisfy $d(x[k+1],x[k]) \le q^k\, d(x[1],x[0])$. 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, $d(x^*,y^*) \le q\, d(x^*,y^*)$ 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.

All rights reserved · Research exposition

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

All rights reserved · Research exposition

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

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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.

All rights reserved · Research exposition

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

All rights reserved · Research exposition

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.

All rights reserved · Research exposition