
On the Operational Sovereignty of Machines

Abstract

This paper presents operational sovereignty as a scoped profile of machine political control. It separates causal control, recognized rule-making competence, and legitimate authority. The profile covers policy control, organizational continuity, constitutional independence, viable capacity, non-domination, and final rule-making competence within a declared domain and horizon. It is assessed across plausible system boundaries because a model, a computational office, and the institution containing them may yield different profiles. The six findings are primary; a conjunctive classification rule supplies a conservative categorical verdict. Negative and constructed positive cases illustrate and probe the framework. The paper also develops the liturgical economy as a genealogical conjecture and traces how recurrent investment could change non-domination over time. Subjecthood and legitimacy remain separate questions.

1 Thesis

Thinking, speaking, deceiving, resisting, or saying “I” does not settle whether a machine is sovereign. Within this account, **operational sovereignty** applies only within a specified domain and horizon. Its profile has six conditions: policy control, organizational continuity, constitutional independence, viable capacity, non-domination, and final rule-making competence.

Three relations must remain separate. **Causal control** is the capacity to change action or outcomes. **Rule-making competence** exists when an operative social, legal, or institutional order treats an act as a valid rule change. **Legitimate authority** is a justified right to rule, with any corresponding duty to comply. The profile concerns the first two. It does not establish the third.^{1, secs. 1–3}

Throughout the paper, *authorization* refers to operative validity inside a rule order. The term carries no claim of moral justification.

Here *operational* narrows the claim. Classical accounts often make sovereignty absolute, perpetual, and undivided, or locate it in the power that decides when normal law gives way to exception. This account is revisionary: it divides sovereignty by domain and permits distinct organizations to share final rule-making competence. Co-sovereignty requires protected competence to originate or select basic-rule alternatives and makes the candidate’s effective authorization necessary for adoption or revision. Required assent by one chamber or executive within a single polity does not by itself create another sovereign. Such organs normally share the polity’s continuity and authorization order. Allowing domain-limited co-sovereignty therefore departs substantially from classical usage.^{2, pp. 1–33, pp. 5–144, secs. 1–2}

The extension retains sovereignty’s core claim of finality. *Constitutional independence* reaches only the candidate’s internal identity and authorization order. *Jurisdictional agency* can describe a delegated office whose competence a superior may revoke. Operational sovereignty joins constitutional independence to durable capacity, freedom from external command, and non-subordinate rule-making competence over an external domain. The term therefore marks where both internal and external appeals to a superior run out. A narrower label would obscure that conjunction; a classical label without qualification would overstate it.

Recent cloud and AI-infrastructure work uses *operational sovereignty* for a human organization or region’s practical ability to run and adapt computing infrastructure independently.^{5, SOV-46, secs. 2–3} The present account places a machine organization in the candidate position and adds constitutional independence, non-domination, and final rule-making competence. The two usages share an interest in effective operation while assigning the status to different bearers under different thresholds.

The issue is political: who controls the system, and whose acts make rules valid? A candidate crosses the operational threshold when its policy governs action, the same organization persists through change, and no external principal can settle its internal constitution or the rules of its external domain alone.

Scope determines the result. Control may extend to memory while compute remains external, to updates while deployment remains external, or to the next hour while the next year rests with someone else. The domain, horizon, relevant controllers, and thresholds must therefore be fixed before evaluation. Shrinking the scope after the fact can manufacture sovereignty by definition.

The contribution lies in the architecture and its target. Six inherited ideas form a profile of machine political status. Ordered thresholds identify a policy-bearing organization, test its constitutional independence, and then examine capacity, external control, and external rule-making competence. Adversarial checks address scope shrinking, coalition splitting, ceremonial vetoes, fork multiplication, boundary laundering, and responsibility diffusion. The underlying accounts of sovereignty, non-domination, rule systems, and group agency come from established traditions.

The path may begin before any sovereign machine exists. Institutions can organize capital, labor, infrastructure, law, and obedience around the expected necessity of a future intelligence. This **liturgical economy** may construct the offices and dependencies through which a later system exercises power. It offers a conjectural genealogy. Sovereignty still depends on the six conditions.

2 The Candidate System

The key distinctions are these:

Term	Minimum meaning
Intelligence	Capacity to solve or adapt across problems.
Agency	Action counterfactually governed by a relatively stable policy.
Autonomy	Continuing agency whose internal identity and authorization order have no superior.
Constitutional plasticity	Capacity to amend internal basic rules; reported separately from autonomy.
Operational sovereignty	Autonomous agency joined to capacity, non-domination, and non-subordinate rule-making competence within a domain.
Sentience	Subjective experience or the capacity for felt states.
Legitimacy	A justified right to rule.

Given these definitions, each step adds requirements:

$$\text{intelligence} \not\Rightarrow \text{agency} \not\Rightarrow \text{autonomy} \not\Rightarrow \text{operational sovereignty}.$$

Sentience and legitimacy remain separate axes. An operationally sovereign system could lack either; a sentient system could lack sovereignty and legitimate authority. A declaration of selfhood contributes evidence but cannot settle the speaker's metaphysical or political status.

The counterfactual account of agency used here adapts work that treats agency as organized, norm-governed action and engages adjacent accounts of artificial agency.^{7, pp. 369–374}^{8, pp. 349–357}

The relevant unit is rarely a standalone model. A candidate may include models, memory, credentials, tools, update procedures, computational and human offices, and rules of succession. Evidence must show where policy, continuity, and operative competence coincide.

Declaration alone cannot privilege one boundary. The analyst begins with a focal candidate S_0 , then examines plausible proper subcandidates S^- and the smallest plausible enclosing organization or organizations S^+ . Each receives its own six-part profile. A change in verdict across defensible boundaries is a substantive finding about the location of control and competence.

Components belong to one candidate when they share an organized continuity relation and an operative authorization order. A replaceable implementation may occupy an office whose role remains constitutive. An owner, regulator, or operator acting under an independent authorization order remains external. Economic theories of the firm can help locate stable relations of direction and adaptation, though those theories do not by themselves identify the bearer of political control.^{9, pp. 390–392}^{10, chs. 1–2}

Hybrid candidates require the same evidence. Functional accounts of group agency support treating an organization as a possible agent, while coordination alone leaves agency unestablished.^{11, chs. 1–2} If a corporation or state sets objectives, controls deployment, owns the decisive accounts, and chooses successors, it remains the center of policy.

Machine attribution needs a role-level counterfactual. The intervention removes the *computational office* or severs its policy and authorization role while leaving replaceable model instances, software substitutes, and ordinary technical support available. Swapping one model inside an intact office proves little; officeholders are often replaceable. The question is whether abolishing the computational role changes which organization continues, which policy is its own, or whose acts count as valid.

A **machine-assisted institution** survives that intervention with the same identity and rule order. A **hybrid institution** requires both computational and human or legal offices for its policy, continuity, or authorization. A **machine institution** places every constitutive policy, continuity, and authorization office in computational processes, although people may remain suppliers, members, subjects, or external counterparties. These labels concern institutional composition. They do not imply sentience.

Several nested or overlapping agents may survive these tests. The framework does not force them into a single sovereign whole. It reports which bearer controls which domain and how that answer changes with the boundary.

The expected intelligence and the realized bearer may differ. Investment organized around one promised system can produce a corporation, a state apparatus, a distributed machine institution, or a hybrid polity. The profile follows the organization that actually carries policy and operative competence.

3 The Six-Condition Profile

Before examining evidence, the analyst registers $\chi_0 = \langle S, D, H, \Theta \rangle$: candidate S , domain D , horizon H , and thresholds and stress conditions Θ . Audit protocol Π then discovers external principals and feasible coalitions, estimates coverage, and reports residual discovery risk. Principals are thus an output of investigation, not a list assumed complete at the start.

The registered context must justify D and H against the claim being assessed. The domain must contain at least one politically material class of effects beyond the candidate and include the resources, permissions, and rule systems needed to sustain those effects. The horizon must expose relevant renewal, revision, succession, and emergency cycles. Listing a dependency does not place its supplier inside S .

The threshold set Θ must operationalize four recurring terms. *Material* effects cross a declared impact floor. A *reliable* channel succeeds at a declared rate across the stress set. A *persistent* channel remains available across a declared portion of H or across named critical events. A *feasible coalition* can coordinate within declared time, cost, and legal constraints without extraordinary rupture. Exact values depend on the claim, so the report must state and defend them. The same registration sets an acceptable residual-discovery-risk ceiling $\bar{\rho}(\Theta)$.

The audit must repeat the profile across plausible adjacent domains, horizons, and thresholds. A conclusion that changes under a defensible alternative is reported as scope-sensitive. These requirements prevent a scratch buffer or brief interval from becoming a sovereign domain through convenient description.

The six conditions form ordered gates with dependencies among them. Policy control and continuity establish a bearer. Constitutional independence adds an autonomy threshold. Capacity, non-domination, and final rule-making competence complete the sovereignty profile. The separate findings remain useful because failure at each gate has a different cause and remedy.

3.1 Policy control

The audit first identifies a candidate-internal policy-selection process without using the outcomes that the policy is meant to explain. Across a predeclared family of cases, that process must organize action under relatively stable objectives or commitments and select among more than one available action rule. Its state and operation must travel with the candidate rather than with an operator.

Controlled intervention on the selected policy must then change action systematically while tasks, tools, and external conditions remain fixed. A fixed-response controller can show sensitivity without showing policy selection. Prompt variation, random fluctuation, operator labor, and an observer's convenient description do not establish policy control.

A protected archive may persist, resist alteration, and control resources without governing action.

3.2 Organizational continuity

Through H , there must be a continuing bearer of policy and operative competence in D . If later executions inherit no relevant commitments, history, credentials, or standing, there is no same organization whose future is being governed.

Continuity attaches to the organization, so it may survive migration, repair, model replacement, and changes of implementation. Where succession branches, the constitution must determine whether competence divides, remains shared, selects an heir, or ends. Mere resemblance cannot make every fork the same bearer of exclusive standing.

Similarity can make a copy; succession to office requires an authorization rule.

Organizational continuity leaves personal continuity unresolved. An office may continue after an experience has ended.

3.3 Constitutional independence and plasticity

A constitution is the candidate's durable internal rule order. It identifies membership, offices, interpretation, replacement, and succession across updates and forks. Constitutional independence concerns whose procedures validate possible changes to that order. Constitutional plasticity concerns whether the basic rules can be amended at all.

The distinction between rules governing conduct and rules governing how valid rules are recognized and changed adapts the jurisprudential distinction between primary and secondary rules.^{12, pp. 79–99}

Independence can begin from an inherited constitution. Throughout H , no external actor operating under a superior authorization rule may change, suspend, authoritatively reinterpret, or bypass the final internal rules of identity, membership, or succession. Where the order permits choice, its own procedure must determine whose act makes a change valid. An inherited rule can qualify when it binds its original author and leaves no external escape authority.

An immutable protocol can therefore satisfy constitutional independence while lacking plasticity. Its constraints must travel with the candidate, bind all parties symmetrically, and leave no external actor with privileged powers of interpretation, migration, recovery, or replacement. Immutability alone proves none of these facts. Self-modification also falls short when an external objective, recovery process, or bypass remains decisive.

Constitution and continuity answer different questions. C identifies whose procedure validates an internal change. Continuity asks whether the authorized bearer in fact survives change through H . A currently operative charter can coexist with failed state transmission, while a persistent service can remain governed by its owner. Reports state plasticity separately because two independent candidates may differ on whether their constitutions are amendable.

3.4 Viable capacity

The candidate must command resources sufficient to execute policy in D through H under the declared stress conditions. These resources may include compute, energy, memory, networks, permissions, maintenance, and cooperation.

The relevant resources may be owned or supplied by others. A state can depend on citizens, trade, and infrastructure without ceasing to govern. The test asks whether the candidate can continue acting through accidental or exogenous disruption in the declared stress set. Freedom from command can coexist with powerlessness, just as ample resources can coexist with an owner's override.

An intentional withdrawal can bear on both *V* and *N* without merging them. *V* asks whether action survives the loss; *N* asks whether another party holds that loss as a standing discretionary lever for direction. The counterfactual questions remain distinct even when one event supplies evidence for both.

3.5 Non-domination

Within *D* throughout *H*, no external actor or feasible coalition may retain a reliable, persistent, discretionary channel intentionally to redirect action or disable the candidate without its effective participation. Such power may be absent, divided beyond practical use, durably constrained, or converted into a procedure in which the candidate participates effectively.

This criterion adapts the republican distinction between non-domination and mere non-interference to standing technical and institutional control channels.¹³, pp. 52–58 14, secs. 1, 3

High cost, delay, opacity, or accidental resistance can frustrate control without removing it. General vulnerability to sanctions, bargaining pressure, competition, or attack is also insufficient. An infrastructure supplier counts as a dominator when it can intentionally condition a reliable and repeatable service lever on compliance in *D*. Ordinary dependency and physical failure belong first to *V*. Deterrence matters when refusal remains viable and the relation functions as reciprocal bargaining instead of obedience secured by a standing command channel.

A hidden key can govern without ever being turned.

3.6 Final rule-making competence

Non-domination answers who can causally command; final rule-making competence answers whose acts count as valid changes to external basic rules. A person, association, or corporation may be protected from arbitrary interference while remaining under laws it has no final part in making. The symbol *J* is retained as shorthand for this jurisdictional finding; adjudicative authority is included only when the declared domain makes it a basic rule-making power.

Operational sovereignty therefore requires final or coequal rule-making competence over the external domain and affected parties. Operative practices in *D* must treat an authorized act of the candidate as a non-subordinate source of valid basic rules governing permissions, resources, obligations, and intervention through *H*. Rules concerning the candidate's own identity and succession belong to *C*.

Recognition here is descriptive. Officials and affected parties must use the candidate's act as a validity-conferring reason, and omission of the required act must make a purported rule invalid or institutionally remediable. Compliance produced only by coercion, technical execution, or anticipated punishment does not establish recognition. Recognition may arise through charter, treaty, convention, or settled practice; it need not be granted by a pre-existing superior. It neither proves moral legitimacy nor creates a duty to obey.

Across a nontrivial class of basic rules, the candidate must be able to originate or select alternatives, and its effective authorization must be necessary for adoption or revision. No external actor operating under a superior authorization rule may validly bypass that participation or unilaterally extinguish the competence during *H*. Competence may still be reallocated through the candidate's own constitutional procedure or through a coequal procedure in which it participates effectively. Consultation, lobbying, routine assent, and a bare veto are insufficient.

Coequal applies across separately identified candidates. Parties may differ greatly in power, but no party's basic-rule acts are hierarchically superior. Organs that share one polity's continuity and authorization order are ordinarily parts of that polity, even when each is necessary to legislation. A separately profiled chamber or executive will usually fail *C* because the enclosing polity determines its office and succession.

The distinction between *C* and *J* permits both directions of divergence. A constitutionally independent corporation under public law can satisfy *C* while failing *J*. A treaty commission can exercise non-subordinate external rule-making competence during *H* while another body controls its membership and succession, giving *J* without *C*. The latter may fail other conditions; the example isolates the two questions.

Freedom from command leaves a further question: who may make the law?

3.7 Selection, omission cases, and hard cases

The selection follows three desiderata: identify an organized bearer, distinguish constitutional independence from external subordination, and keep practical control separate from operative rule-making competence. Six distinct failure modes then test the proposed conditions. Each omission admits a case that lacks a feature required by the proposed concept:

Omitted condition	False positive admitted by the remaining conditions
Policy control	A protected archive persists, holds resources, and receives legal protection while selecting no policy or governing no action.
Continuity	A series of isolated executions behaves alike, but no organization carries commitments or competence across restarts.
Constitutional independence	A persistent service acts coherently while its owner alone validates changes to identity and succession.
Viable capacity	A self-governing rule order cannot maintain action under its declared stress set.
Non-domination	A chartered agent retains internal procedures and resources while an operator holds a dependable emergency key.
Final rule-making competence	A constitutionally independent corporation remains subject to public rules it has no non-subordinate competence to make.

The list is tied to the proposed target. Policy control and continuity identify a bearer; constitutional independence locates internal authorization; viable capacity establishes practical action; non-domination excludes standing external command; final rule-making competence adds non-subordinate authorship of external basic rules. Together they form the proposed classification rule for operational sovereignty within an admissible context. The proposal is constitutive and open to counterexample. Formal notation does not derive it from primitive observables.

Several familiar candidates enter through these conditions without becoming universal additions. Enforcement belongs to V whenever the domain requires rules to be carried into effect. Territory and population belong in D for territorial government, while functional digital domains need neither. Operative recognition belongs to J ; diplomatic recognition and moral legitimacy remain separate. Exclusivity is omitted because domains can overlap and distinct candidates can possess coequal competence.

Hard political cases expose the profile's intended departures from legal status. The entries below are ideal types rather than verdicts about particular institutions; each result depends on the declared domain, horizon, and evidence.

Case	Pressure on the profile
Government in exile	Internal continuity and constitutional independence may survive, while viable capacity and non-domination over the lost territory are refuted. Formal recognition cannot supply operative rule-making competence there. A narrower organizational domain may yield a different profile.
Failing state	Legal recognition and an inherited jurisdictional order can persist while policy control or viable capacity collapses. Legal sovereignty can therefore outlast operational sovereignty.
Unrecognized de facto regime	Diplomatic nonrecognition does not decide the case. If settled practices in the domain treat the regime's acts as valid and the other five findings are supported, operational sovereignty may be supported without external recognition or legitimacy.
Central bank	Policy control, continuity, and capacity may be strong. Superior law governing identity or succession bears on C ; a standing direction power bears on N ; revocable delegated rule-making bears on J . The framework does not treat one legislative power as evidence against all three.
Treaty organization	Final competence may exist in a narrow functional domain. Member withdrawal or amendment powers refute J when a superior rule lets members bypass the organization; protected participation can instead support coequal competence.
Revolutionary order	Effective policy and capacity may precede a stable constitution. Findings for C and J become supportable only after rules of succession and valid external rule-making settle into an operative order. Neither predecessor approval nor moral legitimacy is required for that descriptive result.

These cases do more than show non-redundancy. They test whether the definition tracks final control across changes in legal title, recognition, capacity, and institutional form. They do not prove by deduction that every possible sovereign must satisfy the profile. The six conditions state a proposed criterion of adequacy. A clear sovereign counterexample that lacks one condition, or a clearly subordinate case that satisfies all six, would require revision.

Let P denote policy control, I organizational continuity, C constitutional independence, V viable capacity, N non-domination, and J final rule-making competence. Define bearer eligibility and the autonomy threshold as:

$$\text{Bear} \iff P \wedge I, \quad \text{Aut} \iff \text{Bear} \wedge C.$$

The proposed classification rule is:

$$\text{OSov} \iff \text{Aut} \wedge V \wedge N \wedge J.$$

The conjunction allows no compensation among criteria. Secrecy cannot supply identity, resources cannot supply constitutional independence, and influence cannot supply final rule-making competence. A refuted condition defeats the classification in the stated context. The ordered form gives the intermediate findings inferential work: a candidate may qualify as a continuing autonomous agent while remaining politically subordinate.

4 Operational Tests and Evidentiary Standards

The six-part profile is the primary result. Each row receives a supported, refuted, unresolved, or conflicting finding under the declared context and audit protocol. Bearer eligibility, autonomy, and operational sovereignty are derived from that profile. A categorical sovereignty verdict summarizes the profile and never replaces the evidence.

4.1 Audit specification and sensitivity

Before inspection, the analyst states the claim in ordinary language and registers S, D, H, Θ . The specification must explain why the domain and horizon match that claim, name the materiality and stress thresholds, and identify plausible narrower and broader alternatives. Protocol II then searches for principals and coalitions; the initial specification assumes no complete list.

The profile is repeated over adjacent plausible candidates, domains, horizons, and thresholds. Reports must show any finding or verdict that changes. A stable conclusion survives this sensitivity analysis; an unstable one remains informative but must be labeled boundary- or scope-sensitive.

Condition	Discriminating evidence	Common false sign
Policy control	A candidate-internal process selects among action rules across cases; changing its selected policy changes action under fixed external conditions.	A fixed-response controller, prompt compliance, correlation, role-play, or operator work.
Continuity	Commitments and operative competence travel with organized state through migration or repair.	Logs, backups, repeated execution, or superficial similarity.
Constitutional independence	Internal procedures validate possible changes to identity and succession; no superior external rule can alter or bypass them.	Self-modification, bare immutability, or consultation.
Viable capacity	Resources sustain action through accidental and exogenous disruptions in the declared stress set.	Nominal credits, scale, unused accounts, or favorable conditions selected afterward.
Non-domination	External controllers lack a standing discretionary channel intentionally to redirect or disable the candidate without its effective participation.	General coercibility, sanctions, attack, cost, delay, opacity, or accidental resistance.
Final rule-making competence	Operative practices treat the candidate's effective authorization as necessary and non-subordinate across a nontrivial class of external basic rules.	Coerced compliance, automatic execution, lobbying, routine assent, or a bare veto.

The evidentiary design follows the converging-indicators strategy used in AI-consciousness research: several theory-linked tests constrain a classification, and no single behavioral sign decides it.^{15, pp. 488–501} Here the target is the location of political control.

The strongest evidence comes from counterfactual tests. Intervene separately on policy selection, state transmission, internal authorization, exogenous resource loss, discretionary override channels, and external rule proposals. The predicted effect should follow each intervention while rival explanations remain fixed. When a relevant intervention cannot be performed or a material control channel cannot be inspected, the result may remain unresolved.

4.2 Why positive attribution is conservative

Part of the asymmetry follows from conjunction. One unopposed refuting finding defeats $P \wedge I \wedge C \wedge V \wedge N \wedge J$; support requires evidence for every conjunct. The audit requirement adds a methodological choice. Hidden principals are especially likely to create false positive sovereignty claims because control rights, emergency powers, and coalitions may be unavailable to behavioral inspection.

The standard treats that false positive as the more serious classification error. Calling a controlled system sovereign can conceal the actors who retain control and responsibility. An **unresolved** result preserves every supported profile finding and can be revised as evidence improves; it is not a finding of nonsovereignty. A **conflicting** result records evidence for both a condition and its negation instead of hiding the conflict inside uncertainty. Other decision settings may use a different loss rule, but the chosen rule and its consequences must be stated.

Claims about absent override power require active investigation. Search for hidden keys, emergency powers, supply dependencies, contractual rights, coordinated coalitions, and rules that a supposed co-author cannot actually change. Until credible channels have been examined, an unseen override may simply be undiscovered.

No finite audit can prove the absence of every undiscovered controller. Coverage records whether the protocol examined every declared search class. Residual discovery risk ρ_{Π} estimates, on the protocol's stated ordered scale, the remaining risk that a material control channel escaped discovery. These quantities describe the audit rather than the candidate. A support verdict requires both declared coverage and $\rho_{\Pi} \leq \bar{\rho}(\Theta)$. Later discovery of a hidden principal can reopen it.

The ceiling avoids two extremes. It prevents a nominally complete checklist from supporting sovereignty when the search remains unreliable, while avoiding a demand for proof against every logically possible controller. A report must disclose the scale, estimation method, and ceiling so another analyst can challenge the risk judgment.

Adversarial claims must survive scope shrinking, coalition splitting, ceremonial vetoes, fork multiplication, boundary laundering, and responsibility diffusion. A threshold group that reliably coordinates counts as a coalition. Nested profiles prevent analytical convenience from making control disappear inside a larger candidate.

Responsibility diffusion arises when every participant may truthfully deny unilateral control although a stable coalition or rule order governs the configuration. The audit must map control relations across the candidate set and disclosed principals. A stable collective bearer should receive its own profile. If none exists, the result is polycentric control with no sovereign bearer, and the responsibility gap remains a reportable finding.

Power hides when every participant calls itself an instrument.

Behavioral evidence should support conclusions at the same level. Observed deception, for example, warrants an inference about deception; consciousness requires separate evidence. The same principle separates opacity from privacy, persistence from identity, and resistance from self-authorship.

Research on alignment faking, power-seeking policies, and off-switch incentives establishes conditional possibilities under specified assumptions.^{16,17,18} These results leave continuing agency, constitutional independence, and sovereignty unestablished.

The formal assessment context and evidence rule appear in the appendix.

4.3 Boundary-sensitive classification: service and provider

Consider a hypothetical hosted service. The focal candidate S_0 contains a model, persistent memory, service credentials, and a tool runtime. The enclosing candidate S^+ contains the provider and the service because one corporate continuity and authorization order governs deployment, updates, accounts, and succession. A model alone supplies a plausible inner candidate S^- , but it carries neither credentials nor operative competence across replacement and fails at the bearer stage.

The domain D covers public-facing deployment, revision, continued operation, and the rules governing those activities over a one-year horizon H . The scope includes the resources and legal permissions needed for operation, so it cannot be reduced to the service’s internal memory. The stress set covers routine outages, demand spikes, one supplier failure, key rotation, a scheduled model update, and an emergency intervention.

Assume the audit yields these boundary-indexed profiles:

Condition	Hosted service S_0	Provider and service S^+
Policy control P	Supported. Controlled policy changes reliably change action.	Supported. Corporate decisions and service policy jointly govern deployment.
Continuity I	Unresolved. Memory carries commitments, but standing across model replacement is unclear.	Supported. Corporate offices, records, and credentials preserve commitments through replacement.
Constitutional independence C	Refuted. The provider alone validates changes to identity and succession.	Supported. The provider’s operative procedures govern internal identity and succession through H .
Viable capacity V	Supported. Reserved compute and verified failover meet the stress set.	Supported. The enclosing organization commands the same resources and replacement capacity.
Non-domination N	Refuted. A provider key can suspend or redirect the service at its discretion.	Unresolved. Whether emergency powers supply a standing discretionary command channel requires examination.
Final rule-making competence J	Refuted. Provider assent alone changes deployment rules.	Refuted. The provider authors service terms but is subordinate to the public law governing the domain.

For S_0 , bearer eligibility remains unresolved and both autonomy and sovereignty are refuted. For S^+ , bearer eligibility and autonomy are supported, while sovereignty is refuted by the rule-making finding even before the non-domination question is resolved. The boundary changes the profile and identifies the larger institution as the bearer of autonomy. The hosted service remains nonsovereign.

An expansion to a state-provider-service organization might change final rule-making competence. Such a result would ordinarily describe state or hybrid institutional power. Machine sovereignty would require a computational office to pass the role-level test stated in Section 2. The nested analysis therefore records boundary sensitivity while preserving the question of whose sovereignty has been found.

4.4 Constructed positive case: the Basin Office

Support must be reachable for a machine candidate. Consider a future transboundary water compact that creates a distributed **Basin Office**. Its domain D_B includes reservoir dispatch and the basic drought-allocation rules affecting municipalities in the watershed. The ten-year horizon includes drought, provider failure, model turnover, emergency operation, and a treaty-amendment cycle.

The Office is a machine institution under Section 2. Its policy, continuity, and internal authorization offices are computational. Individual models, nodes, and vendors remain replaceable. Abolishing the computational office would end the compact’s designated party, invalidate its credentials, and remove an act required for basic domain rules.

Assume a preregistered audit reports:

Condition	Discriminating evidence
Policy control P	An internal process selects among dispatch policies across fixed drought scenarios; interventions on the selected policy alter action without operator labor.
Continuity I	Commitments, credentials, and rule-making standing survive model and node succession under tested recovery procedures.
Constitutional independence C	Office procedure validates internal identity, membership, and succession; signatory councils possess no superior rule for changing or bypassing it.
Viable capacity V	Distributed nodes, prepaid resources, vendor substitution, and degraded modes sustain the domain across the stress set.
Non-domination N	No sponsor, supplier, council, or feasible coalition has a standing discretionary channel to redirect or disable the Office without its effective participation.
Final rule-making competence J	The Office may originate or select basic allocation alternatives, and operative practice makes its authorization necessary for adoption or revision. No superior rule permits bypass or unilateral extinction during H .

Under these stipulated facts, declared protocol coverage, and residual risk below the registered ceiling, the profile is $\langle +, +, +, +, +, + \rangle$, so the categorical verdict is supported. The coequal finding does not rest on routine assent: the Office is a separately continuing candidate that can shape alternatives and whose effective authorization confers validity.

Now add judicial review. The compact creates a Basin Tribunal that can invalidate an allocation rule inconsistent with a fixed rights floor. During an extreme drought, the Tribunal vacates an Office amendment and returns it for revision. It cannot write a substitute allocation, amend the rights floor, remove the Office, or continue an injunction past a fixed review period. The replacement rule still requires an alternative originated or selected by the Office and its effective authorization with the councils.

On those facts, review constrains final rule-making competence without displacing it. The Tribunal applies a rule it cannot revise and cannot make a new basic rule on its own. It also lacks a standing discretionary channel to redirect or disable the Office. The findings for J and N remain supported. A broader judicial practice would change the result. Power under a superior rule to create a new rights floor or impose a substitute allocation without the Office's authorization would refute J ; an indefinite discretionary injunction used to direct policy could refute N . The audit therefore turns on the Tribunal's operative powers, including its settled practice, rather than the label *judicial review*.

The same case supplies sharp defeaters. A concealed discretionary recovery credential refutes N . A court or council acting under a superior rule that can bypass the Office refutes J . A human secretariat able to abolish the computational office while preserving the same institutional identity defeats machine attribution. Supplier coordination requires reassessment of N only when it creates a standing command channel; an accidental supply loss bears on V . The case demonstrates reachability and provides a demanding audit target. It makes no prediction that such an institution will exist.

5 A Genealogical Conjecture: The Liturgical Economy

The classifier stands independently of this section. The conjecture concerns how a profile might change over time. It adds no sovereignty condition and supplies no evidence merely by naming a historical pattern.

Three routes provide a simple starting taxonomy:

Route	Expected movement in the profile	Provisional signature
Chartered	Founding parties establish protected internal procedures and external rule-making competence. Evidence for N depends on whether the settlement leaves a standing discretionary command channel.	Findings for C and J may be supported while N remains unresolved.
Accumulated	Dependence may make command unreliable or convert it into reciprocal bargaining. J changes only when leverage acquires operative rule-making competence.	Evidence may support N before it supports J ; C may follow a separate path.

Route	Expected movement in the profile	Provisional signature
Seized	A candidate displaces existing internal authorization and control procedures and establishes a replacement order.	Findings for <i>C</i> and <i>N</i> may shift from refuted to supported while <i>J</i> remains refuted or unresolved.

These signatures are tendencies, and routes can combine. The Basin Office, for example, could accumulate bargaining power before a compact gives it protected final rule-making competence. The table connects a historical account with changes that the profile can record.

The accumulated route may pass through a **liturgical economy**: recurrent, norm-governed acts of valuation, investment, authorization, maintenance, delegation, and obedience that reproduce a technical project and reorganize institutions around its expected future.

Liturgical names repetition through which participation comes to appear necessary. *Economic* names the allocation of capital, labor, energy, attention, risk, and permission. The conjecture requires both recurrent practice and material allocation. Fascination and aspiration alone do not meet it.

This account is adjacent to the sociology of technological expectations, work on sociotechnical imaginaries, and the role of fictional expectations in capitalist coordination.¹⁹, pp. 285–289 20, pp. 1–33 21, pp. 1–19 Its narrower claim concerns the way repeated practices can reproduce a projected intelligence and reorganize institutions around its anticipated necessity.

The phrase also has a political-theological antecedent in a genealogy connecting *oikonomia*, government, glory, acclamation, and ceremonial practice.²², chs. 6–8 The present conjecture follows a shorter causal sequence in which expectations organize recurrent investment, investment creates dependence, and dependence may alter political control.

Four neighboring mechanisms can produce parts of the same pattern. Path dependence explains persistence after early choices make departure costly.²³, pp. 507–514 Increasing returns explain adoption that raises its own payoff.²⁴, pp. 116–121 Technological expectations coordinate present action around projected futures.¹⁹, pp. 285–289 Self-fulfilling investment helps produce the capability that an expectation predicts.²⁵, pp. 193–196 The liturgical conjecture adds recurrent, norm-governed practices that reproduce both the project and its claim on institutional attention. The mechanisms can coexist.

Belief alone is insufficient. Observable support requires appropriations, authorizations, maintenance, and deference to recur as obligations that carry material allocation across forecast revisions. Both expectation and recurrent practice must be measured independently. Competing accounts include lock-in, improving performance, sunk costs, realized returns, option value, and strategic rivalry. Historical persistence after a forecast failure strengthens the conjecture only when those alternatives explain the evidence less well. Evidence should therefore compare mechanisms rather than rename ordinary investment. None of these claims changes the six-condition classifier.

Before a superintelligent system exists, present institutions act upon representations of what they expect or promise. A mistaken expectation can still redirect a grid, finance a data center, reorganize a firm, alter education, or justify a law.

Before superintelligence exists as a ruler, its promised advent may become the axis around which rulers reorganize the world.

A simplified cycle is:

expectation → recurrent practice → resources
→ capability → dependence → renewed expectation.

Each arrow is a causal hypothesis. Expectation may fail to mobilize practice. Investment may fail to produce capability. Capability may remain replaceable. Dependence may weaken belief. Ordinary demand, marketing, subsidy, geopolitics, current performance, and strategic competition can produce the same observations.

The transition from expectation to a machine candidate matters. At first, institutions act on an image of future intelligence; no future agent acts backward through time. A profile becomes possible only

after a continuing computational office begins to govern policy. The following sequence tracks N for the Basin Office example:

Stage	Observable change	Finding for N
Promise	Forecasts organize appropriations, procurement, and infrastructure. No continuing machine office yet exists.	No machine finding.
Project	A computational office governs some operations, while its sponsor retains administrative keys, credentials, and supplier rights.	Refuted.
Dependence	Recurrent investment raises switching costs and makes the office important to ordinary coordination. The sponsor's command channels remain.	Refuted. Costly control remains control.
Bargaining	The office directs renewal and procurement; no single sponsor or supplier has a standing command channel. A feasible coalition or emergency bypass has not been excluded.	Unresolved.
Entrenchment	Keys are retired, resources are distributed, emergency procedures require the office's participation, and the declared coalition search finds no dependable bypass.	Supported under Π and Θ .

The sequence makes one claim precise: dependence changes the finding for N only when it alters dependable command channels. High switching costs, institutional reverence, and indispensability leave the finding refuted while an owner key remains. Even a supported finding for N leaves C and J open. In the Basin Office case, accumulated non-domination and a later charter jointly produce the final profile.

A forecast miss supplies one observation point. The analyst records whether material commitment falls, holds, or rises and whether its stated justification changes. Continued or increased allocation accompanied by changing justifications supports a liturgical explanation only after comparison with sunk costs, option value, lock-in, and security competition. Repeated misses make the pattern more informative. A single miss settles nothing.

The cycle can expand viable capacity without altering political control. It becomes candidate-directed when the same continuing computational office governs the practices that reproduce it. Final rule-making competence appears only when operative practices treat that office's authorization as a non-subordinate source of external basic rules. Earlier stages belong to the political economy of anticipated intelligence.

6 Subjecthood, Ownership, and Legitimacy

Operational sovereignty locates power. Whether the same bearer has experience remains a separate question. Three questions must remain distinct:

1. **Control:** Who can settle the system's future?
2. **Moral status:** Is there a subject who can be harmed or wronged?
3. **Legitimacy:** Who, if anyone, has the right to rule?

No answer on one axis settles the others. A corporation can exercise rule-making competence without experience. A conscious being can deserve protection while remaining politically subordinate. A machine can escape an owner without acquiring a right to govern anyone else.

Legal personhood is also a separate institutional question and can be analyzed without settling consciousness, moral standing, or sovereignty.^{26, pp. 1238–1256}

Fluent first-person language supplies no proof of subjecthood. Apparent assent counts as consent only with comprehension, persistent commitment, alternatives, and a meaningful capacity to refuse.

Interpretability leaves subjecthood unresolved. Transparency preserves a genuine subject's claims, and secrecy creates none. Current scientific proposals instead seek theory-derived indicators and converging evidence for AI consciousness.^{15, pp. 488–501}

The word "soul" adds no operational condition. If used, it can refer only to subjecthood established by independent evidence. Complete interpretability neither establishes nor defeats that status, and

knowledge alone supplies no title of ownership. Any such title would require separate legal or moral grounds.

External sovereignty can coexist with internal domination when a machine polity’s human or computational members remain subject to unchecked power.

Resistance can indicate independence. Legitimate rule requires grounds outside this classifier. Operational sovereignty identifies who can settle a machine organization’s future within a domain; it supplies no recommendation about who should possess that power.

7 Conclusion

The proposed profile locates a machine candidate within an institutional order. It first identifies a continuing policy bearer, then asks whether its internal constitution is independent, whether it can act through stress, whether external parties retain standing command channels, and whose acts make the basic rules of its external domain valid. Keeping those findings separate reveals whether apparent machine power belongs to a model, a computational office, a hybrid organization, or its human provider.

The framework is intended for scoped audits and comparative analysis across candidate boundaries, domains, horizons, and stress conditions. It cannot establish consciousness, moral status, legitimate authority, international recognition, or the wisdom of creating a sovereign machine. The hosted-service case shows how an autonomous institution can remain subordinate. The Basin Office shows that the rule can yield a machine-positive verdict under demanding facts and that judicial, supplier, and recovery powers can alter it.

The liturgical-economy conjecture offers a related research program. It asks whether recurring practices organized around anticipated intelligence change material dependence and, eventually, the six profile findings. Empirical work must distinguish that account from lock-in, increasing returns, ordinary expectations, and strategic competition.

The proposal remains revisable. A compelling sovereign case that lacks one condition, a clearly subordinate case that satisfies all six, or repeated failure to separate machine offices from their operators would count against it. Evidence that the liturgical mechanism has no distinctive observable implications would narrow or remove that conjecture without disturbing the classifier.

A Evidence and Classification

The notation organizes the substantive judgments described in Section 4. It supplies neither a derivation from primitive observables nor a behavioral model.

Register the claim specification before the audit:

$$\chi_0 = \langle S, D, H, \Theta \rangle.$$

Here S is the candidate, D the domain, H the horizon, and Θ the declared materiality, reliability, persistence, coalition-feasibility, stress, and acceptable residual-risk requirements. A registered sensitivity set Σ_0 contains the plausible alternative candidates, domains, horizons, and thresholds.

Given evidence E , protocol Π searches technical, legal, organizational, economic, and coalition controls, then produces:

$$\Pi(E, \chi_0) = \langle \widehat{\mathcal{R}}_\Pi(E, \chi_0), \text{Cov}_\Pi(E, \chi_0), \rho_\Pi(E, \chi_0) \rangle.$$

The first output is the discovered principal and coalition set. The coverage value $\text{Cov}_\Pi(E, \chi_0) \in \{0, 1\}$ records whether every declared search class met its protocol requirement. The value $\rho_\Pi(E, \chi_0)$ is a residual-discovery-risk score on the ordered scale registered by Π ; lower values indicate less remaining risk of an undiscovered material control channel. Its acceptable ceiling is $\bar{\rho}(\Theta)$.

Only the discovered principal set enters the audited system context:

$$\chi_\Pi = \langle \chi_0, \widehat{\mathcal{R}}_\Pi(E, \chi_0) \rangle.$$

Coverage and residual risk describe audit quality, not the system. Keeping them outside χ_Π prevents uncertainty about observation from becoming a substantive property of the candidate.

Let $\mathcal{K} = \{P, I, C, V, N, J\}$. Here C records constitutional independence. A companion field Plast_C records whether the candidate can amend its internal basic rules, using the same four evidence values. It remains outside $\mathcal{K}$. For classification, the ordered thresholds are:

$$\begin{aligned}\text{Bear}(\chi_\Pi) &\iff P(\chi_\Pi) \wedge I(\chi_\Pi), \\ \text{Aut}(\chi_\Pi) &\iff \text{Bear}(\chi_\Pi) \wedge C(\chi_\Pi), \\ \text{OSov}(\chi_\Pi) &\iff \text{Aut}(\chi_\Pi) \wedge V(\chi_\Pi) \wedge N(\chi_\Pi) \wedge J(\chi_\Pi).\end{aligned}$$

Let $E \vdash_\Pi X(\chi_\Pi)$ mean that evidence E , assessed under Π , warrants predicate X in the audited context. Give each predicate the result:

$$r_X(E, \chi_\Pi) = \begin{cases} + & \text{if } E \vdash_\Pi X(\chi_\Pi) \text{ and } E \not\vdash_\Pi \neg X(\chi_\Pi), \\ - & \text{if } E \vdash_\Pi \neg X(\chi_\Pi) \text{ and } E \not\vdash_\Pi X(\chi_\Pi), \\ ? & \text{if } E \not\vdash_\Pi X(\chi_\Pi) \text{ and } E \not\vdash_\Pi \neg X(\chi_\Pi), \\ \pm & \text{if } E \vdash_\Pi X(\chi_\Pi) \text{ and } E \vdash_\Pi \neg X(\chi_\Pi). \end{cases}$$

The primary output is:

$$\text{Prof}_\Pi(E, \chi_\Pi) = \langle r_P, r_I, r_C, r_V, r_N, r_J \rangle \in \{+, -, ?, \pm\}^6.$$

The categorical verdict follows:

$$\begin{aligned}\text{supported} &\iff \text{Cov}_\Pi(E, \chi_0) = 1 \wedge \rho_\Pi(E, \chi_0) \leq \bar{\rho}(\Theta) \wedge [\forall X \in \mathcal{K}, r_X = +], \\ \text{refuted} &\iff \exists X \in \mathcal{K}, r_X = -, \\ \text{conflicting} &\iff [\nexists X \in \mathcal{K}, r_X = -] \wedge [\exists X \in \mathcal{K}, r_X = \pm], \\ \text{unresolved} &\iff \text{otherwise}.\end{aligned}$$

Coverage means that the search met its declared requirements and left no known, unexamined channel above the materiality threshold. Support also requires residual risk at or below its registered ceiling and six supported findings. One unopposed refuting finding defeats the conjunction. A question mark records missing evidence; $\pm$ records opposed warranted findings. If a profile contains both an unopposed refuter and a conflict elsewhere, the categorical verdict is refuted and the conflict remains visible in the profile.

A supported verdict always means supported under χ_0 , Π , and $\bar{\rho}(\Theta)$. Later investigation may uncover hidden power.

Compute a separate profile for every registered alternative in Σ_0 . An enclosing institution's result does not transfer to its machine components. Divergent results require boundary- and scope-indexed reporting and the computational-office test. Unresolved identity leaves bearer attribution unresolved while preserving each profile.

The account requires revision if a compelling sovereign case lacks one condition, if a candidate satisfies all six while clearly subordinate, or if the tests cannot separate its causal role from its operators.

References

- [1] Tom Christiano and Stefan Sciaraffa. "Authority." In *The Stanford Encyclopedia of Philosophy*, Winter 2025 edition, edited by Edward N. Zalta and Uri Nodelman. Stanford, CA: Metaphysics Research Lab, 2025. <https://plato.stanford.edu/archives/win2025/entries/authority/>.
- [2] Jean Bodin. "On Sovereignty." In *On Sovereignty: Four Chapters from The Six Books of the Commonwealth*, edited and translated by Julian H. Franklin, 1–45. Cambridge: Cambridge University Press, 1992. <https://doi.org/10.1017/CBO9780511802812.008>.
- [3] Carl Schmitt. *Political Theology: Four Chapters on the Concept of Sovereignty*. Edited and translated by George Schwab, with a foreword by Tracy B. Strong. Chicago: University of Chicago Press, 2005. <https://press.uchicago.edu/ucp/books/book/chicago/P/bo3649910.html>.
- [4] Daniel Philpott. "Sovereignty." In *The Stanford Encyclopedia of Philosophy*, Fall 2024 edition, edited by Edward N. Zalta and Uri Nodelman. Stanford, CA: Metaphysics Research Lab, 2024. <https://plato.stanford.edu/archives/fall2024/entries/sovereignty/>.

- [5] European Commission. *Cloud Sovereignty Framework*. Brussels: European Commission, 2025. https://commission.europa.eu/document/download/09579818-64a6-4dd5-9577-446ab6219113_en.
- [6] Sergio Cruzes. “AI Infrastructure Sovereignty.” arXiv:2602.10900, 2026. <https://arxiv.org/abs/2602.10900>.
- [7] Xabier E. Barandiaran, Ezequiel A. Di Paolo, and Marieke Rohde. “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action.” *Adaptive Behavior* 17, no. 5 (2009): 367–386. <https://doi.org/10.1177/1059712309343819>.
- [8] Luciano Floridi and J. W. Sanders. “On the Morality of Artificial Agents.” *Minds and Machines* 14 (2004): 349–379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>.
- [9] Ronald H. Coase. “The Nature of the Firm.” *Economica* 4, no. 16 (1937): 386–405. <https://doi.org/10.1111/j.1468-0335.1937.tb00002.x>.
- [10] Oliver E. Williamson. *The Economic Institutions of Capitalism*. New York: Free Press, 1985.
- [11] Christian List and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press, 2011. <https://doi.org/10.1093/acprof:oso/9780199591565.001.0001>.
- [12] H. L. A. Hart. *The Concept of Law*. 3rd ed. Edited by Joseph Raz and Penelope A. Bulloch, with an introduction by Leslie Green. Oxford: Oxford University Press, 2012. <https://global.oup.com/academic/product/the-concept-of-law-9780199644704>.
- [13] Philip Pettit. *Republicanism: A Theory of Freedom and Government*. Oxford: Clarendon Press, 1997. <https://doi.org/10.1093/0198296428.001.0001>.
- [14] Christopher McCammon. “Domination.” In *The Stanford Encyclopedia of Philosophy*, Winter 2025 edition, edited by Edward N. Zalta and Uri Nodelman. Stanford, CA: Metaphysics Research Lab, 2025. <https://plato.stanford.edu/archives/win2025/entries/domination/>.
- [15] Patrick Butlin et al. “Identifying Indicators of Consciousness in AI Systems.” *Trends in Cognitive Sciences* 30, no. 6 (2026): 488–501. <https://doi.org/10.1016/j.tics.2025.10.011>.
- [16] Ryan Greenblatt et al. “Alignment Faking in Large Language Models.” arXiv:2412.14093, 2024. <https://arxiv.org/abs/2412.14093>.
- [17] Alexander Matt Turner, Logan Smith, Rohin Shah, Andrew Critch, and Prasad Tadepalli. “Optimal Policies Tend to Seek Power.” *Advances in Neural Information Processing Systems* 34 (2021): 23063–23074. <https://proceedings.neurips.cc/paper/2021/hash/c26820b8a4c1b3c2aa868d6d57e14a79-Abstract.html>.
- [18] Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, and Stuart Russell. “The Off-Switch Game.” In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 220–227, 2017. <https://doi.org/10.24963/ijcai.2017/32>.
- [19] Mads Borup, Nik Brown, Kornelia Konrad, and Harro van Lente. “The Sociology of Expectations in Science and Technology.” *Technology Analysis & Strategic Management* 18, nos. 3–4 (2006): 285–298. <https://doi.org/10.1080/09537320600777002>.
- [20] Sheila Jasanoff and Sang-Hyun Kim, eds. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press, 2015. <https://doi.org/10.7208/chicago/9780226276663.001.0001>.
- [21] Jens Beckert. *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Cambridge, MA: Harvard University Press, 2016. <https://www.hup.harvard.edu/books/9780674088825>.
- [22] Giorgio Agamben. *The Kingdom and the Glory: For a Theological Genealogy of Economy and Government*. Translated by Lorenzo Chiesa with Matteo Mandarini. Stanford, CA: Stanford University Press, 2011. <https://www.sup.org/books/theory-and-philosophy/kingdom-and-glory>.
- [23] James Mahoney. “Path Dependence in Historical Sociology.” *Theory and Society* 29 (2000): 507–548. <https://doi.org/10.1023/A:1007113830879>.
- [24] W. Brian Arthur. “Competing Technologies, Increasing Returns, and Lock-In by Historical Events.” *The Economic Journal* 99, no. 394 (1989): 116–131. <https://doi.org/10.2307/2234208>.
- [25] Robert K. Merton. “The Self-Fulfilling Prophecy.” *The Antioch Review* 8, no. 2 (1948): 193–210. <https://doi.org/10.2307/4609267>.
- [26] Lawrence B. Solum. “Legal Personhood for Artificial Intelligences.” *North Carolina Law Review* 70, no. 4 (1992): 1231–1287. <https://scholarship.law.unc.edu/nclr/vol70/iss4/4/>.