The institutional world model

Continual learning across sovereign virtual machines

Raeez Lorgat
Momentum Research · research.momentum.inc
raeez@momentum.inc

21 September 2025

Abstract

AI agents conducting institutional work need an environment that determines how law, regulation and policy apply to proposed actions. New cases and changing fraud methods require learning from rulings, investigations and outcomes. We introduce an institutional world model: a learned representation of rule application and institutional consequences. Standing mandates authorize qualifying updates, while each sovereign environment controls which interpretations govern its agents. Learning crosses institutional boundaries with zero-knowledge proofs, subject to receiving adoption and use.

We construct a finite consent learner whose proofs establish a fit over a complete source-attested inventory and inference for an exact request. Adoption and current-use checks connect the calculations to an institutional effect. Public weights and decisions can reveal private inputs. Finite results address correction and continuing duties, since completion under fixed models need not survive switching. A synthetic lifecycle follows private learning through registration, correction and payment after operator replacement. A separate fraud comparison finds harmful direct transfer. Receiving validation discards all source weights and matches local learning without incremental transfer benefit. Improvement against adapting adversaries remains an empirical question.

1. Introduction

An AI agent conducting business acts through institutions whose rules determine what an operation can do. Payments use accounts, contracts create obligations, and transfers change ownership. Before allowing these effects, the environment must determine how its governing requirements apply to the case. A rule's words alone may leave questions about control, consent or evidence.

Consider an agent asking to change a company's registered owner. Some changes require additional consent. A fixed check may miss that the buyer and seller share a controller. Competent rulings give examples of when consent is required; investigations can reveal conduct that defeats a control or legitimate activity it obstructs. An unchanged procedure may continue missing a newly identified pattern or obstructing legitimate activity.

Continual learning must connect new evidence to institutional decisions. A fitted model alone cannot supply that connection: predicting official approval confers no official power, and a fraud score establishes no adjudicated finding. The institution must authorize a use, reserve appropriate judgments and provide for correction. A receiver may also have different rules, case meanings and populations, while source cases remain private.

We introduce the **institutional world model**, a learned representation of rule application and institutional outcomes within a sovereign execution environment. Governing material supplies requirements and rulings teach their application; observations and investigations support empirical models of behavior and consequences. An adopted procedure uses these distinct

sources to assess what an agent may do. We specify the cycle from evidence to candidate, adoption, execution and correction.

Mass is a decentralized network of sovereign execution environments for AI agents. Each environment learns the correct, safe interpretation of its sovereign law, policy and regulation, so it keeps up with fraud and bad actors on its own. The companion network paper develops the execution architecture [1]. Each environment is a **sovereign virtual machine**, governed by an institution that supplies its rules, delegated powers and authoritative state. Here we develop the learning process that changes its operative interpretation, including how other environments can use what it learns.

A standing mandate authorizes an update procedure in advance. The environment checks each candidate against that mandate, current sources and work underway; qualifying updates can be adopted automatically. Zero-knowledge proofs verify specified learning and inference without transmitting raw case values, although outputs can still reveal them. The receiver checks local meaning, validation evidence and authority for use. Institutions can learn from one another while retaining different operative interpretations.

Adoption also affects work underway. Registration may have occurred while payment remains pending. Correction must retain that registration occurred and account for payment the bank can still perform. A stricter model can remove the only permitted route for an existing refund. Model updates must therefore account for continuing duties, original deadlines and possible external effects. Replacing a model, agent or operator does not restart these accounts.

We study two finite tasks. Synthetic consent rulings distinguish ordinary transfers, common-control reorganizations and nominee or custody arrangements. The procedure requires consent, reserves judgment or satisfies that requirement, leaving other authority, eligibility and resource checks intact. A fraud task compares local learning with direct and locally validated transfer. Admission and investigation select later evidence, and other actors respond to controls. The fixed comparison establishes no success against adapting adversaries.

Our contributions are a private fitting and inference construction, a model-adoption contract, and finite results for correction and continuation through updates, under stated source, cryptographic, mediation and service premises. Sections 2–3 connect evidence to an exact action; §§4–5 develop receiving use and private verification. Sections 6–8 treat updates, progress, correction and succession. Section 9 gives the evaluations, and §§10–11 discuss related work and conclusions. The appendix supplies the restricted admission algebra.

2. Evidence and its limits

The learning cycle begins with a choice of evidence. A ruling, an investigation and the absence of an observed loss support different conclusions. We first specify what each record means, then ask which questions the available records can answer.

2.1. Rulings, findings and missing outcomes

A consent label answers a particular authority's question about a particular rule and case. The learning record must identify that authority, the source revision and the permitted purpose of the determination. A reversal changes the current label. The earlier ruling and its effects remain in the record.

Empirical labels need comparable precision. Suspicion, referral, investigation, competent finding and reversal record different events. Refusing a proposal supplies no fraud finding about it. Missing or censored outcomes remain explicit, with the investigation that could establish them.

The observation contract records how the environment obtained each label, with its proposal, acting capacity, affected objects, operative policy, decision, observation horizon and available

outcome. The task determines needed relationships among entities, controllers, documents, transfers and their times. The access contract determines which it may use.

Identity errors can attach rulings or payment observations to the wrong case. Source assertions therefore remain separate from current assessments. SamePerson identifies records about one person without merging estates. SameEvent identifies records about one occurrence, so corrections can change payment support without creating another payment. Training binds admitted label revisions and the source cut defining available versions and observations.

Authenticity cannot recover an outcome that was never observed. Suppose a gate refuses 100 of 1,000 proposed transactions and permits 900. Nine permitted transactions later receive confirmed fraud labels. Assume complete labels for those 900 at the declared horizon. Also stipulate fixed binary potential outcomes under permission, without cross-proposal interference or policy-induced response changes.

If none of the refused transactions would have been fraudulent, permitting all proposals would yield a fraud rate of 0.9%. If every refused transaction would have been fraudulent, the rate would be 10.9%. The observed history is identical in both cases.

Proposition 1 (Unidentified outcomes and interpretations). Under the fixed-outcome premises, a record containing only the admitted outcomes in this example identifies the all-proposals potential fraud rate no more narrowly than $[0.009, 0.109]$. More generally, two admissibility functions that agree on every recorded case but disagree on a new program cannot be distinguished from that history alone.

Proof. Assign each missing potential outcome zero or one. Both extreme assignments preserve every observed label and decision. They attain the stated endpoints. For the second claim, the two interpretations supply identical learner inputs and different required answers. A function of those inputs cannot return the required answer in both worlds. Randomisation cannot guarantee both answers either. □

This is the selective-label obstruction [2]. An additional investigation, permitted experiment or justified identification assumption can distinguish some of the possible worlds. With interference, the counterfactual model must also cover changed outcomes among formerly admitted proposals. Provenance identifies the supplied observations. It cannot resolve what they leave open.

The same limitation applies to interpretation. Two procedures can fit every retained ruling and disagree on the next ownership request. Selecting between them requires a normative source, authoritative examples or a competent determination. Repeating the already observed cases need not provide that information.

2.2. Observation sufficiency

Even an observed fact can become unavailable through the choice of representation. Consent may depend on who controls whom. A summary containing only earlier verdicts can discard the relationship needed for the next request. Empirical transfer analysis faces the same problem.

Consider four entities and three transfers. The histories $A \to B \to C \to A$ and $A \to B \to C \to D$ can have identical transaction counts and verdicts. Only the first contains a directed cycle. The cycle supplies a feature for investigation. It does not establish fraud.

Let $T(H)$ be the retained representation of history H, and let Q be a declared family of questions.

Proposition 2 (Observation sufficiency). Every $q \in Q$ has an exact decoder from $T(H)$ precisely when

$$T(H) = T(H') \quad \Longrightarrow \quad q(H) = q(H') \qquad (q \in Q).$$

Proof. Equal decoder inputs must have equal outputs. Conversely, the condition assigns one answer to each fibre of T, defining a decoder on its image. □

Proposition 2 identifies the questions a representation preserves. Implementation also needs an available decoder. The result requires no unlimited retention. When a later task needs discarded relationships, the institution must obtain source evidence, change its observation contract or report that it cannot answer.

Some local learning therefore needs personal or entity-level records. A model exported without rows can still retain information about those records. Access, retention and release obligations follow the task and its artifacts. Section 5 distinguishes the information hidden by a private calculation proof from information disclosed by its outputs.

3. Learned interpretation and admission

3.1. The authority of learned outputs

The evidence limits determine what a learned output can establish. Consent rulings concern a rule's application, while fraud findings support predictions about behavior. Before assigning either output a consequence, the institution must specify the proposition on which it intends to rely.

Role	Output	Basis for reliance
Empirical model	Prediction, uncertainty or identified range	Observation and statistical contract
Institutional interpreter	Application of a named rule to a case	Competent adoption, precedents and reserved decisions
Admission policy	Permitted effect, hold, question or refusal	Current authority and complete execution contract

One neural model may perform all three roles in the table. Each interface must establish the proposition required for its use. A prediction of official approval confers no power, and a high fraud score establishes no adjudicated finding. The institution may nevertheless adopt a score as grounds for a defined protective measure.

For example, an adopted threshold can trigger a bounded investigation hold. The authorizing procedure states the affected action, duration, review and challenge route. Those terms give the score its operational force. They supply no implicit power to confiscate assets, make a criminal finding or interfere with unrelated remedial work.

Competent adoption determines which judgments the institution reserves and which it delegates to learned assessment. A standing mandate can permit automatic updates within that boundary. Competence over every interpretation a model can predict is not presumed.

3.2. Finite consent learning

We use a finite consent task to connect an exact fitting procedure to its adopted consequence. A supported feature extractor assigns each labeled case to a cell x. Within that cell, parameter $\theta_x \in (0,1)$ predicts the binary recorded outcome. Let n_x count observed labels and k_x the positives. The finite learner minimizes

$$L_x(\theta_x) = -(k_x + 1)\log\theta_x - (n_x - k_x + 1)\log(1 - \theta_x).$$

Lemma 3 (Finite fitted parameters). The unique minimiser is

$$\hat{\theta}_x = \frac{k_x + 1}{n_x + 2}.$$

Proof. The derivative vanishes exactly when $(k_x+1)(1-\theta_x) = (n_x-k_x+1)\theta_x$. The second derivative is positive throughout $(0,1)$, and the loss diverges at either endpoint. □

The added units are declared pseudocounts. They define the estimator without supplying missing outcomes or establishing accuracy. With no observed labels, a cell returns $1/2$, and the local training record identifies the lack of support. A private-weight receiver cannot distinguish that cell from one with balanced labels by its score alone. The consent procedure therefore refers every $1/2$ cell. It does not claim that the receiver knows the sample count.

The consent source supplies 24 synthetic determinations, eight in each category. A positive label records the source authority's determination that the named rule required additional consent.

Source category	Determinations	Consent required	Fitted score	Adopted response
Ordinary transfer	8	0	1/10	No additional consent
Common control	8	7	4/5	Require consent
Nominee or custody	8	4	1/2	Refer

The adopted procedure returns NoAdditionalConsent at scores at most $1/4$, RequireConsent at scores at least $3/4$, and Refer between them. An unsupported cell thus receives Refer. This synthetic task stipulates both the category definitions and the competent source's assignment of cases to them.

The fitted score summarizes the admitted rulings. The adopted procedure gives that summary a consequence for the next request. NoAdditionalConsent satisfies only the additional-consent requirement. RequireConsent calls for exact scoped consent, and Refer calls for the reserved determination. The transfer's other admission checks remain necessary.

Six competent reversals of common-control positives reduce the positive count from seven to one, giving a fitted score of $1/5$. RequireConsent becomes NoAdditionalConsent. A suspended case can use that change only after current reassessment and reconciliation. Section 9.2 applies the same estimator to empirical fraud findings under a separately adopted consequence policy.

The fit record identifies the objective, admitted labels, excluded missing labels, feature semantics, data cut and parameters. These bindings are necessary even for this small estimator. A richer neural learner changes the inference and assurance obligations, while the institution still determines the output's meaning and permitted use.

3.3. The exact action and its continuation

The fitted consent response answers only one requirement. We must connect it to the complete assessment of the request before allowing an institutional effect. For a specified request, context and use, an adopted procedure can return

$$\mathsf{Permit}(e, k, v),$$
$$\mathsf{Hold}(b, q, t),$$
$$\mathsf{Refuse}(b, u).$$

In Permit, e identifies the effect and v the evidence supporting it. The continuation k records unfinished work and the conditions for resumption. Hold identifies the basis b, pending question q and deadline t. Refuse identifies its basis and scope u. The operative procedure determines when each result is available.

The host controls protected institutional state. For the ownership request, it checks current power, applicable requirements, exact parties and effect, resources, dependencies and residual duties. Sanctions evaluation is one possible requirement. Complete mediation requires every protected change to pass admission, including administrative and model-adoption operations [3].

Admission must also allow work that a decision creates or leaves pending. A transfer can create a reporting duty due thirty days later. Requiring the report first would obstruct the act that creates the duty. An unresolved application may still permit evidence submission, identity correction or payment of an existing refund. Permission therefore attaches to an exact action, rather than a permanent favorable status for the entity.

Recording the decision supplies no missing permission. Complete mediation can use several physical functions or writers. A small verifier can check the complete admission record, including reasons, unresolved questions and continuing duties.

Appendix A gives a restricted algebra for combining requirement grades. Performing a filing, obtaining consent or discharging a duty still requires evidence and an authorized operation.

3.4. Operative bundles and update mandates

Fixed weights do not fix behavior. A changed feature extractor can move a case to another cell, and changed calibration or thresholds can give its score another consequence. Institution i therefore adopts the complete operative bundle

$$B_{i,v} = (\theta, F, C, P, R, U; E),$$

Component	What it fixes
θ	Learned parameters
F	Feature extraction
C	Interpretation of scores and uncertainty
P	Mapping from interpreted outputs to consequences
R	Inference semantics and reproducibility
U	Permitted uses
E	Training cuts, derivation parents and assurance evidence

The bundle belongs to the governed context Γ_i.

In the private consent profile, the source attests each case's category encoding. The proof checks the calculation over that encoding. Proving feature extraction from committed documents would require a richer profile and its own source and semantic obligations. The finite construction makes no such claim.

Update authority belongs to a separate protected mandate M_i. It identifies the competent origin, updater, baseline, allowed changes, checker, horizon and cumulative accounts. A candidate can reference that mandate but cannot replace its authority. A successor mandate needs competence from the predecessor context.

An assessment binds the actual bundle and its inputs: source and retrieval cuts, extracted features, case, purpose and inference output. The runtime computes the supported inference or verifies a proof of that exact relation. Before commitment, the host rechecks dependencies and current authority. Changed preprocessing, weights, calibration, thresholds or retrieved material require reassessment. A checked correspondence may instead justify reuse for the exact present use. We call that reuse assessment transport.

A joint operation binds the relevant vector of local contexts and bundles. Institutions may train locally, share parameters, delegate training, use local adapters or retain a frozen model. Local control concerns the adopted interpretation. It does not require each sovereign VM to train a separate neural network.

4. Learning across sovereign environments

4.1. The source result and the receiving use

The local procedure gives the fitted model an authorized role. A receiving registrar must determine its own role for the same artifact because its rules may treat those relationships differently. The source may offer a ruling, a learned interpretation or a procedure for acting on that interpretation. Recognizing one ruling does not adopt a general model, and adopting a consent model does not authorize unrelated refusals.

The receiving use has three contracts. Representational correspondence binds features, labels, capacities, time windows and source meanings. Statistical assurance identifies the evidence for performance in the receiving population. Institutional recognition determines which consequence the receiver may give the result. All three are directional and task-specific.

For consent, a correctly fitted source weight can still encode a relationship that receiving law treats differently. The receiver can adopt a correspondence, reserve the question for its own authority or decline the interpretation. The source can continue using its own model.

Recognition can also stop at the evidence. A receiver might request more documents where the source requires consent, or adopt local calibration under an existing mandate. Shared learning is compatible with different operative interpretations.

4.2. Population and consequence

An empirical model can preserve its meaning and still perform differently after transfer. Take a detector with sensitivity 0.9, false-positive rate 0.01 and prevalence π. Its positive predictive value is

$$\frac{0.9\pi}{0.9\pi + 0.01(1-\pi)}.$$

At prevalence 0.1, the value is 10/11. At prevalence 0.001, it falls to 10/121. Among 100,000 destination proposals, the latter population gives ninety flagged fraudulent cases and 999 flagged benign cases. The calculation holds classifier rates fixed. It estimates no country's rates.

The unchanged threshold therefore imposes a different burden. Receiving-population evidence must address both performance and the costs of the adopted response. Causal transportability needs additional assumptions about differences between source and destination [5].

An embedding can represent jurisdictions numerically through features of their legal traditions. It may provide a prior for investigating transfer, but supplies no transfer theorem. Weight deltas also need compatible parameterization and semantics. Without that compatibility, the

exchanged object must be a complete detector, rule proposal or evidence packet with a specified receiving interpretation.

4.3. Verification before adoption

Before adoption, the receiver authenticates the inventory and verifies the calculation against expectations in protected policy. It checks correspondence, validation and proposed use under its existing authority. Each institution controls its interpretation, so shared parameters require no global synchronization.

Authentic, correctly computed models can use inaccurate labels or perform poorly at their destination. Model-replacement attacks in federated learning can preserve ordinary performance while introducing targeted errors [6]. The receiving mandate specifies the required checks. Finite validation can reject the authenticated poisoned candidate tested in §9, without excluding every unseen backdoor.

Failures must remain separately contestable. Corrected source findings may require retraining, while changed receiving law may require a different use of a correct model. A computation proof establishes neither population fitness nor adoption.

5. Private learning and inference

5.1. Source inventory and commitments

The receiving procedure requires evidence of the fit, but the source's cases may be private. We construct a proof of the specified calculation over a complete finite inventory, or **manifest**, attested by the source. The manifest identifies case revisions, commitments and the learning task. It lets the receiver check the calculation without receiving raw case values.

A commitment hides an encoded value and binds the source to it. For each row, six hidden indicators represent the three categories paired with the two possible rulings.

Use the Ristretto group of prime order

$$\ell = 2^{252} + 27742317777372353535851937790883648493.$$

Fix generators G, H whose discrete-logarithm relation is unknown. The source commits to six values for each row i, one for each category $j \in \{0, 1, 2\}$ and label $b \in \{0, 1\}$:

$$C_{ijb} = z_{ijb}G + r_{ijb}H.$$

The private z_{ijb} values select the row's category and label. Independent uniform scalar blindings r_{ijb} hide their openings. A labeled row has exactly one indicator equal to one. A row with a missing label has six zero indicators. The public manifest records label presence, censoring and observation time. Its case identifiers are unique and sorted.

Let p_i be the public indicator that a label is present. Let e_i indicate that the label is present, uncensored and available by the training cut. The relation requires

$$z_{ijb} \in \{0, 1\}, \qquad \sum_{j,b} z_{ijb} = p_i, \qquad n_j = \sum_i e_i(z_{ij0} + z_{ij1}), \qquad k_j = \sum_i e_i z_{ij1}.$$

The manifest and public eligibility rule determine every contributing row. Commitments supply its private category and label values. An inventory contains between one and 256 rows. The source attests the census and encoding under the named rule. That attestation supplies the coverage premise beyond the arithmetic checked by the proof.

The artifact identifies the task, normative source, category contract, learner, runtime and fixed decision policy. It supplies three canonical fractions a_j/b_j, with $0 < a_j < b_j \leq 258$. It binds the training cut, manifest, recipient, purpose, permitted actions and release identity. The

correction frontier gives the latest admitted correction sequence and head hash. A historical replacement link can identify an earlier artifact without asserting continued computational reliance on its invalidated examples.

The source signs under a distinct protocol domain. Canonical encoding and content hashes bind the manifest's exact bytes. Protected local policy supplies the receiver's expected source key, task and current manifest frontier. Descriptions supplied by the caller cannot establish those expectations.

The complete inventory prevents selective inclusion within the manifest. Proving that chosen rows belong to a collection would still allow unfavorable rows to be omitted. Here the verifier derives the relation from every signed row and the public eligibility rule. Coverage beyond that manifest remains the source's responsibility.

5.2. The training proof

An aggregated Bulletproof checks verifier-derived commitments in eight-bit ranges [13]. For each indicator commitment C, the verifier includes C and its complement commitment $G - C$. For each row it derives

$$R_i = \sum_{j,b} C_{ijb} - p_i G$$

and includes the residual commitment R_i and its negative $-R_i$.

For category j, the fitting equation is

$$a_j(n_j + 2) = b_j(k_j + 1).$$

The verifier derives commitments to the counts and fitting residual:

$$\begin{aligned} C_{n_j} &= \sum_i e_i(C_{ij0} + C_{ij1}), \\ C_{k_j} &= \sum_i e_i C_{ij1}, \\ F_j &= a_j C_{n_j} - b_j C_{k_j} + (2a_j - b_j)G. \end{aligned}$$

Both F_j and $-F_j$ enter the same proof, with openings equal to the fitting residual and its negative. The verifier derives them from the manifest. The prover therefore cannot substitute independently chosen counts.

Each row contributes twelve bit/complement entries and two residual entries. The three fitting equations add six, for a total of $14N + 6$. Identity commitments pad this count to a power of two. The transcript binds the full artifact and signed manifest, so inventory, fit and output form one statement.

Proposition 3a (Private training relation). Assume adaptive knowledge soundness and zero knowledge for the instantiated aggregated Bulletproof. Assume an unknown discrete-logarithm relation between the fixed Pedersen generators, Ed25519 authenticity, SHA-256 collision resistance and injective shared encoding. For an accepted envelope outside the corresponding failure events, knowledge extraction supplies openings of every manifest commitment satisfying the stated bit, row and eligibility relations. Its public fractions equal $(k_j+1)/(n_j+2)$ for those openings.

Proof. Knowledge extraction supplies bounded openings for all range entries, including each original commitment. Each bit and its derived complement open to nonnegative integers at most 255. Their sum is one modulo ℓ. Commitment binding and $510 < \ell$ make the integer sum one, so the original value is zero or one.

A residual commitment and its negative also have nonnegative openings at most 255. Their integer sum is zero, so both openings are zero. Binding then identifies each residual with the

verifier's linear combination of the extracted original openings. Row residuals establish the one-hot or all-zero condition. Fitting residuals establish the displayed equation modulo ℓ.

The admitted bounds give $n_j \leq 256$, $k_j \leq n_j$, $a_j \leq 257$ and $b_j \leq 258$. Each product in the fitting residual is at most $257 \cdot 258 = 66{,}306$. Their absolute difference is therefore at most 66,306, below the group order. The modular equality is consequently an integer equality. Positive denominators and canonical reduction give the claimed fractions. Authentication and exact encoding bind those equations to the source's signed inventory and the receiver's expected task. □

Knowledge extraction is essential here. Pedersen commitments are perfectly hiding, so existence of an opening alone would not connect the proof to the claimed calculation. The argument requires extracted openings with the stated algebraic correspondence.

Ganesh et al. prove simulation extractability for Fiat–Shamir Bulletproofs in the programmable random-oracle model [14]. This protocol-family result leaves an instantiation premise for the particular library and transcript used here.

Proposition 3a establishes the fit over the attested inventory. The truth and legal meaning of a label, and the competence to issue it, remain source premises. Generalization beyond that inventory needs separate evidence. Section 4 specifies the receiving conditions, and §6 examines the changing evidence produced by repeated use.

5.3. Inference for the present request

The training proof establishes the fit. A second proof must connect that fit to the receiver's exact request. To conceal the request's category, the case source supplies a signed credential containing three commitments

$$U_j = u_j G + s_j H, \qquad u_j \in \{0, 1\}, \qquad \sum_j u_j = 1.$$

Public fields bind the task, normative source, category contract, case, subject, revision, source cut and request digest. Source competence and correct encoding remain premises, as they do for training.

The adopted fractions fix public decision codes c_j: zero for NoAdditionalConsent, one for RequireConsent and two for Refer. For a claimed code $o \in \{0, 1, 2\}$, the inference proof establishes

$$\sum_j c_j u_j = o.$$

The proof checks each bit and complement, both signs of the one-hot residual and both signs of the decision residual. Ten entries pad to sixteen. Its transcript binds the artifact, signed credential and complete inference statement.

Proposition 3b (Private inference relation). Under Proposition 3a's cryptographic and encoding assumptions, an accepted inference envelope establishes knowledge of a one-hot opening of the signed case credential. The claimed output equals the fixed adopted policy's decision for that category. The statement binds this conclusion to its exact artifact, case revision, request and proposed operation.

Proof. Extract the three committed openings. The bit/complement checks and one-hot residual give exactly one selected category by Proposition 3a's integer argument. The two decision-residual checks give $\sum_j c_j u_j = o$. All decision values lie between zero and two, so no modular ambiguity remains. Signature verification and the bound transcript identify the credential, artifact and statement used in these equations. Exact comparison with the host's expected statement fixes the present request and operation. □

The host derives its expected statement from the protected current context and proposed effect. Separate authenticated ingress establishes the current case credential before inference can authorize use. The host checks that credential, the adopted artifact and any required consent or reserved determination. A proof for another operation fails the exact statement comparison. NoAdditionalConsent satisfies only the additional-consent requirement. Powers, resources and native-effect checks still apply.

Verified inference can thus enter the same assessment relation as local inference. The proof establishes the output. Receiving adoption and current-use checks establish the institutional consequence permitted for that output.

5.4. Disclosure through the output

Both proofs disclose their public statements, which include exact weights, metadata and decisions. Zero knowledge limits additional information from the proof. The statement itself can still reveal a category or label, so private verification needs an explicit release contract.

Take an inventory with one eligible row. The selected category has weight $1/3$ for a negative label and $2/3$ for a positive label. Every other category has weight $1/2$. The published weights identify both category and label. An inference outcome likewise identifies a category when the three categories have distinct public decisions.

The release contract must cover that output. In the finite consent profile, case identifiers, source revisions and other manifest metadata are public. Commitments and proofs avoid publishing raw witnesses. They do not prevent inference of those values from permitted outputs.

Increasing the aggregate size does not by itself remove disclosure. For binary records, $\theta = \sum_{j=1}^{500} y_j/500$ reveals the last label to someone who knows the other 499. A corrected weight θ' can reveal a changed label through $500(\theta - \theta')$. Neither withholding rows nor imposing a population floor supplies a privacy guarantee.

A statistical privacy contract specifies the protected unit, neighboring datasets, bounded contributions, recipients and complete observable transcript. The transcript includes any visible weights, gradients, scores, model selection, correction releases, refusals and timing. Repeated inference can disclose information even when the model remains local. The observer model must include every recipient coalition covered by the claim.

For neighboring histories $H \sim H'$, differential privacy requires

$$\Pr[M(H) \in E] \leq e^{\varepsilon} \Pr[M(H') \in E] + \delta$$

for every observable event E. The network supplement gives finite mechanisms with bounded contributions, complete output channels, durable release identities and cumulative accounting [1]. The institution must adopt a particular mechanism and its allowed outputs. A cohort threshold of 500 can supplement that policy, but proves no privacy property by itself.

Learning only from an already released private statistic is post-processing under the same observation contract [7]. Reading corrected private data generally incurs another release cost. A retry must reuse the retained release identity, sample and bytes. Renaming the model does not renew its allowance. Deleting a training row cannot retract an earlier release.

The singleton example rules out a nontrivial differential-privacy guarantee for the exact-weight profile. Across neighboring singleton labels, the event releasing one of the distinct weights has probabilities one and zero. The inequality then requires $\delta \geq 1$.

Changed source rows and case credentials need fresh independent uniform blindings. Reusing a blinding exposes a changed indicator through $C' - C \in \{-G, 0, G\}$. Fresh blindings prevent that direct comparison. Weight changes and correction metadata remain disclosed.

A proof can also verify a computation over a supplied random tape without establishing how the tape was sampled. Statistical privacy needs that sampling law and an explicit contract for the combined release.

Some required performances themselves disclose protected information. Suppose neighboring worlds require different visible payments, each forbidden in the other world. The event containing one world's required payment has probability zero in the other. The privacy inequality bounds successful payment in the first world by δ.

A stronger proof system cannot alter this bound. Changing the observer, receipt, intermediary or permitted disclosure can change the institutional problem. The authorized change must be explicit and cover any later payment that would reveal the same information.

6. Repeated updates and their accumulated effects

The disclosure account already extends beyond a single proof. The same is true of institutional burdens and delay. We now specify what the update procedure must retain across a sequence of models.

6.1. Cumulative limits

An individual update limit need not bound the sequence's effects. Repeated reductions of at most 0.01 can move a threshold across its entire range. Burdening at most one percent of the population per update can burden everyone if the groups rotate. Changing the model identifier can also conceal the restart of an old hold.

The monitor must therefore persist outside the bundle. Over the mandate's horizon, it retains the authorized baseline and accumulated burdens, costs, disclosures and exposure. Each case keeps its original clock. A distance bound on weights may restrict engineering change. To control an institutional consequence, it also needs a bound on the resulting decisions or effects.

The mandate can distribute powers within the procedure. It may allow automatic changes to evidence routing while reserving final refusal for a competent determination. An emergency mandate may permit bounded blocking with review and expiry. Objectives, loss weights, label meanings and consequences all belong to the delegation.

A standing mandate authorizes the defined sequence of qualifying updates. An explicit adoption authorizes one identified bundle. In either case, the existing context supplies authority before the candidate can govern a later action.

An accepted update installs the complete checked bundle and adoption record atomically. The candidate cannot change the criteria that authorize it. Assessment and commitment bind the installed version.

6.2. Mandate renewal

Under the finite allowance, each adoption consumes one credit. Continued adoption eventually requires a successor mandate authorized in the exact predecessor context.

Renewal declares authorized changes to future action scope, objectives, baselines and expenditure caps. It retains source frontiers, correction status, model ancestry, completed acts and possible old commands. Disclosure, risk and burden already incurred remain spent. Competent renewal may enlarge future allowances, but the accounts must retain prior expenditure and state any changed lifetime guarantee's resulting total. Renaming a mandate restores no privacy or risk allowance.

Existing duties retain their goals, safety conditions, original deadlines and checked certificates for permitted completion routes. Each also retains its separate allowance for model adoptions affecting its completion route. Neither service steps nor renewal of the learner's general budget replenishes this allowance. Otherwise administrative activity can postpone payment indefinitely, even when each model permits completion. Section 7's result requires actual transitions satisfying its common progress condition and service premise.

A competent disposition can change the performer, timing or assurance while identifying the change and retaining the former promise's status. Renewal cannot record an unpaid duty as complete. Implementing successor mandates remains separate from the finite proofs for admitted bundles and counted execution.

6.3. Policy feedback and adversarial response

The next candidate also faces evidence shaped by earlier use. After a control detects abuse, adversaries may change transaction forms, switch intermediaries or manipulate learning inputs. Legitimate applicants may change behavior to avoid delay. Investigations and corrected findings can reveal those responses. Assessing the new candidate must include the burdens already imposed on legitimate work.

Let D_t denote data generated in period t and π_{B_t} the active bundle's admission policy. The observations depend on that policy and the responses it induces:

$$D_t = D(\pi_{B_t}, \Gamma_t, \text{responses of other actors}).$$

Retraining therefore acts on a population that earlier controls may have changed. Additional observations need not improve the model or make its parameters converge.

For example, take $Y \sim \text{Bernoulli}(1-\theta)$ under the deployed predictor. Exact squared-error retraining gives $\theta_{t+1} = 1 - \theta_t$. Starting at 0.2 yields the endless cycle $0.2, 0.8$. Every fit is the population optimum, but the population response prevents convergence.

Performative prediction gives conditions for retraining to converge [4]. Applying them requires evidence about the response map and loss. A count of institutions or successful adoptions cannot establish those conditions. Convergence also needs a separate connection to the institution's objective.

Even without feedback, the historical data may not identify a policy's improvement. Suppose two equally common types historically receive actions zero and one, with respective success rates 20% and 80%. Two models reproduce that record. In one, action one succeeds at 80% for both types. In the other, it succeeds at zero for type zero and 80% for type one.

Always taking action one then gives 80% success in the first model and 40% in the second. Both have the same historical 50% baseline. Choosing between these effects needs additional model assumptions or information from a competently authorized investigation or experiment.

6.4. Fixed evaluation criteria

Feedback makes the evaluation criterion consequential. We must fix the population, outcome, baseline and horizon before comparing procedures. Weakening the consent requirement could raise the immediate pass rate while changing the task. The intended outcome may instead be correctly completed ownership changes, with required consent and documentation by a stated deadline.

Likewise, a formation procedure that accepts preliminary declarations must be assessed through verified documentation at the agreed deadline. First-attempt acceptance leaves that duty unresolved. The comparison must account for abandonment, false refusal, fraud loss, investigation, delay and unfinished duties.

Power to adopt, correct implementation, evidence of improvement and acceptance of the tradeoff are separate judgments. An authorized procedure need not improve the selected outcome. A changed objective must be reported as such. Excluding difficult applicants cannot establish improvement for the original population.

7. Continuation under changing interpretations

Keeping the objective fixed requires checking whether the revised procedure can still perform work already owed. Safety of each newly admitted action is insufficient for that promise.

7.1. Continuation under a fixed bundle

The consent procedure can require permission before a future transfer. A new empirical detector can also affect an existing obligation by removing the payment action needed to discharge it. We distinguish discretionary adoption, which can be refused, from binding authority changes that must take effect.

Suppose a refund is owed in state s, and g records its completed payment. The old bundle permits $s \to g$. A new detector flags the recipient and removes that action. The narrower action set can preserve safety while making the refund impossible.

Fix a finite state set S, a safe set $K \subseteq S$, a goal set $G \subseteq K$ and a declared response model. For each initiated action, $\mathrm{Succ}(s, a)$ contains every complete response path's endpoint, and every intermediate state must also be safe. The response model includes old commands that can still perform.

Let $A_B(s)$ contain the actions currently admitted under bundle B. Define

$$\mathrm{Pre}_B(X) = \{s \in K : \exists a \in A_B(s),\ \emptyset \neq \mathrm{Succ}(s, a) \subseteq X,\ \text{all intermediate states lie in } K\}.$$

Starting with $W_0 = G$, set $W_{n+1} = W_n \cup \mathrm{Pre}_B(W_n)$. The finite fixed point is W_B.

Theorem 4 (Continuation under a fixed bundle). Each state in W_B has a strategy reaching G within at most its first-entry rank in completed response steps. The strategy stays in K, provided actual responses satisfy the declared nonempty response contract and its steps complete.

Proof. Give $s \in W_B$ the least rank $r(s)$ with $s \in W_{r(s)}$. Rank zero means the goal already holds. A state of positive rank entered through an admitted action whose complete responses end in the preceding approximation. Select that action. Every response stays safe and ends at a strictly smaller rank. Induction on the rank gives the bound. □

For the refund, $W_{\text{old}} = \{s, g\}$ and $W_{\text{new}} = \{g\}$. Adoption must address that lost route. It can retain an authorized continuation or record a competent change to the obligation, performer, timing or assurance. An authorized release can end the duty without constituting payment.

The goal, response model and inherited duty identity are protected inputs. A candidate cannot redefine payment as assessment or omit the possible arrival of an old payment. A local timeout cannot free a reserve while the provider can still perform. The response contract must be nonempty and cover intermediate safety. An attractive endpoint cannot substitute for an implementation or conceal unsafe execution.

7.2. Progress through changing models

Now let bundle B_0 permit only $s_0 \to s_1 \to g$, and let B_1 permit only $s_1 \to s_0 \to g$. Either fixed bundle allows completion from both states. Switching after each first step produces $s_0, s_1, s_0, s_1, \ldots$. The refund is never paid.

Each step lowers the active bundle's rank. Changing the bundle changes that rank, so the sequence makes no lasting progress. The continuing duty needs a progress condition that survives adoption.

For a finite construction, fix the goal and a common rank $\rho : S \to \mathbb{N}$. Adoption must preserve the goal and cannot increase the current rank. Each promised service step needs an admitted implementation whose possible responses remain safe and strictly lower the rank. Administrative changes must leave sufficient opportunity for those service steps.

Theorem 5 (Continuation through admitted updates). Assume rank zero implies the fixed goal. Suppose each positive-rank state receives a completed promised service step within at most b admitted transitions, for a fixed positive integer b. If every intervening transition preserves safety and does not increase ρ, and each service step strictly decreases it, the goal occurs within at most $b\rho(s_0)$ admitted transitions.

Proof. Before the first service step, the rank is at most $\rho(s_0)$. That step lowers it by at least one. Repeat over blocks ending at service steps. After at most $\rho(s_0)$ such blocks the rank is zero; each block contains at most b admitted transitions. □

The count includes admitted context, service and recorded environment transitions. It bounds neither elapsed time nor unrecorded events. The promised services must actually complete. An absolute deadline additionally requires an actual clock, with no restart on model adoption or mandate renewal.

A finite game can treat more general cases. Its state includes the bundle, update authority, clocks and unfinished work. Its environment includes permitted model changes and provider responses. The common rank gives the simpler sufficient condition used here.

7.3. Revocation and failed assurances

A binding revocation may remove the only lawful performance action. Once the source contract admits it as effective, affected new use must be fenced immediately. Refusing a discretionary adoption cannot keep withdrawn authority in force.

The institution records the failed assurance alongside the surviving duty, exposure and possible commands. Preserving history or balancing the ledger does not perform the service. Replacing the agent cannot make silence count as payment.

A competent disposition may establish a different continuation contract or discharge the duty through an authorized release. It may change the performer, timing or assurance. The former objective remains unmet unless its specified performance occurred. Necessary institutional change can proceed while leaving its consequences open to correction and remedy.

8. Correction and continuity of the institution

8.1. Correction dependencies

A source ruling can influence a receiving model and suspend an ownership request that never cites the ruling itself. Correcting direct citations alone would leave that request governed by the invalidated basis.

The derivation graph must therefore cover source and label revisions, training cuts, base models, local adapters, calibration, adopted bundles and assessments. An edge records possible dependence. Observed absence of correlation cannot remove a protected dependency. Selective independence needs a checked argument. Conservatively marking every use of an affected bundle is sound.

A delivered correction identifies its authority, exact source revision, scope and operative time. After applying its recognition contract, the receiver marks dependent current reliance for reassessment, justified transport or competent disposition. The correction status remains on source revisions and derivation ancestors. Future derivation and adoption must inherit it or discharge it through checked reconciliation.

The private consent profile binds each signed correction to the previous manifest, correction sequence and exact prior row. The replacement preserves case identity and advances the revision. Unchanged rows stay identical. New rows require separate source admission, and withdrawn rows retain their identities.

The replacement manifest binds the consecutive correction history, and a new proof checks the fit over that manifest. Import uses the admitted current source frontier. An old proof can

remain mathematically valid while failing that currentness check. Adopting the replacement still leaves affected cases to be reconciled.

Case credentials have a separate authenticated history. Admitting a new case revision invalidates earlier reliance even before a replacement inference proof exists. Reconciliation must use the admitted current credential. It cannot supply its own purported current version. Appendix O of the companion supplement specifies both source and case interfaces [1].

The kind of correction determines its retrospective meaning. New evidence, reversed findings, prospective law, population drift and corrected identity require distinct treatment. A fresh fit states its current computational dependencies. Historical replacement links can remain without continuing computational dependence on invalidated labels.

The earlier assessment stays attributable. A replacement model cannot reverse a completed transfer, recall a disclosure, compensate a loss or extinguish an executable bank command. Each consequence needs its own competent act. Correction must reach the institution's economic state as well as its learning record.

8.2. Eligibility after correction

Invalidation must also permit justified resumption. Suppose an identity mistake makes one payment appear twice, causing an allocation check to block new work. A competent SameEvent correction removes the duplicate occurrence. Once the supported allocation passes current checks, the old discrepancy cannot remain an unexplained veto.

An input correction can also change a refund under an already adopted formula. If the current formula gives twenty where earlier evidence supported sixty, the dependent duty needs recomputation or reassessment. An identity authority can correct the input within its powers. Amending the award formula requires the authority competent over that award.

A change beyond the formula's authorized scope needs a competent disposition. Earlier performance and outstanding commands remain in either case. Updating weights alone cannot resolve the changed duty.

8.3. Currentness and receiving authority

A receiver cannot infer a reversal that has not arrived. Consider two worlds with the same receiving state. In one the source finding remains effective. In the other it has been reversed, but notification is delayed. Computation on the identical state returns the same answer in both.

Stronger currentness must come from the source contract. A validity lease, serialized current-use service or bounded notification arrangement can provide it on specified terms. Otherwise correction completeness begins after admitted delivery. The receiver retains stale or uncertain status when a stronger assurance is unavailable.

Receiving information can precede receiving authority. A source reversal may trigger review while the receiver obtains its own competent determination. The local consequence remains indexed by purpose and current receiving rules.

Binding authority changes follow the immediate invalidation rule of §7.3. Reconciliation may remain pending after new use is fenced. Unpaid claims and possible commands survive even if no replacement continuation is yet available.

8.4. Succession and retained evidence

A replacement agent or operator needs the operative bundle, compatible inference implementation, licenses, derivation references, evidence access and pending corrections. Case clocks, mandates, cumulative expenditure and old commands accompany them. Continued use requires available bytes and services, beyond the identity supplied by a content hash.

A private proof may remain verifiable after raw examples disappear. Retraining, challenges or other questions may still require additional evidence. Proposition 2 identifies which questions the retained representation preserves. Authorized erasure can preserve some capabilities and retire others. Requests lacking required evidence must return "evidence unavailable."

A source-label correction states whether the model was retrained, withdrawn, shown independent of the record or remains potentially influenced. Rolling back weights establishes none of these outcomes, and prior disclosures remain recorded.

Writer transfer or a reused process name confers no registrar, tribunal or provider powers. The custody contract tests each promised service's actual post-transfer authority. A successor identity service needs a current scoped appointment or access to an available competent principal.

Independent operation requires usable artifacts and rights, evidence access, available inference, a qualified successor and means to contest earlier decisions. The service contract and fulfillment evidence must cover these dependencies. A writer fence prevents further protected writes by the outgoing operator.

8.5. The combined execution relation

Adoption, use, correction and succession change the same continuing institution. We represent its local state by $X = (\Gamma, H, S, K)$. The context Γ holds rules, powers, active bundles and standing mandates. History H retains attributable events and derivations, while state S holds positions, resources, exposure and persistent accounts. Cases, assessments and unfinished work belong to K.

The following transitions connect learning to protected execution.

Transition	Check at the protected boundary	Result
Generate	Observation access and candidate identity	Candidate and training record, with active context unchanged
Adopt	Predecessor competence, actual bundle and derivation status, mandate monitor and continuing-work disposition	Atomic bundle installation and adoption record
Assess	Exact bundle, inputs, source cut, use and supported inference	Computed result with costs, disclosures and case clock
Use	Current context, dependencies and complete action contract	Exact authorized effect and continuation
Transport	Current authority and checked correspondence for the exact present use	Original assessment retained with its receiving-use witness
Invalidate	Binding authority change under its source contract	Affected new uses fenced, with failed assurance and live claims retained
Correct	Competent correction and conservative dependency closure, including future descendants	Affected current reliance marked or reconciled
Succeed	Accepted custody cut, fenced writer and available post-transfer authority	Continuing model and case operation under the successor

A private artifact enters as a candidate after source and proof checks. Proposition 3a establishes its training relation. Proposition 3b establishes supported inference for an assessment.

The host must bind these statements to the actual current context and effect. Neither proof receipt installs a bundle.

Theorem 6 (Governed learned admission). Start with an admitted local state and the stated transition rules. Assume complete mediation, atomic protected transitions, sound bundle/inference and dependency checks, and the bound source and service contracts. Every finite admitted execution has these properties:

1. Every operative bundle has a predecessor-authorised adoption lineage.
2. Every learned assessment used for an effect binds the current bundle directly, or has a currently valid checked transport witness for that use. Its original computation remains attributable.
3. Model changes retain historical consumption, disclosures and effects. They preserve relevant command possibilities until their effect contract supports exclusion. Competent dispositions can change current positions and future mandate terms without erasing that history.
4. After an admitted correction, no affected current reliance authorises a further effect without the prescribed reconciliation.
5. Imported models have operative force only under the receiving institution's adoption contract.

Proof. Induct on the admitted trace. Generation adds a candidate and cannot alter the active context. Adoption checks authority in its predecessor and installs exactly the checked successor; the persistent monitor supplies the next account state. Assessment computes the supported inference and records its full dependencies. Use compares those dependencies with the current protected state and checks the exact consequence, so a changed model or feature pipeline cannot reuse the old admission without a checked current transport witness. Transport preserves the original computation and establishes the required present correspondence.

Correction first finds the conservative transitive closure. It invalidates or reconciles each affected current reliance while preserving the historical records, disclosures and command identities. The same status checks cover future derivation and adoption, so a stale inactive candidate cannot bypass the correction. The use rule excludes unresolved reliance. Binding invalidation fences new uses immediately and records the assurance failure without declaring the old goal complete. Succession transfers the complete model and case cut under an exclusive writer fence, retaining the competent paths promised by its service contract. Ordinary economic transitions preserve or competently revise positions through the surrounding execution rules. None of these steps turns a foreign artifact into receiving authority. Each case preserves the claimed properties at the next state. □

Theorem 6 is a result about the stated transitions. Implementation must connect bundle installation, inference, database writes and administrative routes to those rules. The bounded public reference requires exact current-bundle equality. General assessment transport needs its own checked correspondence.

The private-proof adapter must likewise obtain its expected statement from protected state and apply the accepted result to the exact operation. A correction need not invalidate the proof's mathematics. Before use, the adapter must compare the bound source and case revisions with their admitted current frontiers.

Theorem 6 alone permits refusal of every proposed act. Theorem 5 supplies progress under its additional continuation and service premises. Useful operation also needs evidence that legitimate work can proceed and that justified refusal preserves the duties left unfinished.

9. Evidence from two finite tasks

The finite tasks test two distinct questions. The consent lifecycle tests whether private learned interpretation reaches the intended effect and remains correctable. The fraud comparison tests whether transferred learning improves the receiving decision under fixed conditions.

9.1. Consent, registration and payment

The 24-determination consent task exercises the construction with actual training and inference proofs. Source meanings and authority are synthetic premises. Section 3.2 gives the exact fitted scores and decisions. Before adoption and use, the receiver checks the complete signed inventory and current case credential against protected expectations.

The initial common-control request requires consent and fails native admission. Six authenticated reversals invalidate its basis before a replacement proof is available. The receiver then verifies the new fit and renews its exhausted adoption allowance under predecessor authority. After adoption, the case remains blocked until explicit reconciliation. That step changes the response to NoAdditionalConsent and retains the original hold clock, deadline and burden.

The host commits one ownership change to Carol, consumes seven service units and records a 100-unit payment command. A seventh source correction arrives after registration while payment is pending. Signed custody carries the command, learned state and pending reconciliation to the successor. A later synthetic provider observation records the full original payment, reducing the outstanding amount to zero. The run reopens the database and compares the complete history with the final state.

Seventeen host tests use real proofs to exercise changed operations, stale source and case frontiers, unauthorized adoption and correction before replacement inference. Renewal preserves lifetime accounts and the unfinished case's update allowance. Public proof packets, accepted statements and the receiving trace supply the evidence. Source-local witnesses remain undisclosed.

The execution was recorded on 24 October 2026. It demonstrates the tested path from private learning to a stored effect under synthetic authority and provider premises. It establishes no general implementation refinement, population accuracy or economic advantage from transfer. The second evaluation measures transfer in a fixed empirical task.

9.2. A fixed fraud comparison

For the empirical task, we use a synthetic Indonesian financial-integrity office and a synthetic receiving office. These constructed roles establish no actual finding, appointment or deployment. The comparison tests how a learned relational pattern affects protected decisions, transfer and correction.

Each proposal contains account-controller relationships, two transfer times and an admitted local-practice record. The feature extractor gives local practice priority. Otherwise it selects a shared-controller and rapid-relay pattern, then the ordinary cell. The practice record provides context without establishing lawfulness.

Lemma 3 gives each cell weight $(k+1)/(n+2)$ from n available binary findings, of which k are positive. Three exact rational weights form the parameter vector. Missing findings are excluded and reported. A score at or above the adopted threshold starts a bounded hold and investigation. It supplies no fraud finding.

Source training, receiving training, validation and test contain distinct cases and immutable label events. Source training has 900 available ordinary findings, nine positive, and 100 available pattern findings, ninety positive. These counts give weights of 5/451 and 91/102. Under an adopted half-threshold procedure, the pattern cell causes a bounded hold and referral.

Source training also includes a local-practice cell and missing findings. A benign receiving practice resembles the source pattern. The protocol fixes every stratum, data cut and missing-label count.

Evaluator truth remains separate from training and model selection. The four arms are rules only, local learning, direct source model and threshold, and receiving-validated transfer. Local learning and receiving-validated transfer have identical calibration capacity.

Both calibrated arms mix initial weights with weights fitted on receiving validation and use the same finite threshold set. Selection minimizes validation cost: eight units for a missed positive and one for a held negative. Brier score, threshold and mixture break ties in that order. The selected policies are evaluated on the untouched test cut.

All arms face the same 320 proposals, each for amount 100, over a five-tick horizon. A permitted proposal executes at tick zero. A hold reserves 100 until investigation at tick two, costing three service units. The investigation perfectly reveals the stipulated evaluator truth, refusing a positive case and executing a negative case. That accuracy is an explicit service assumption.

Every benign case carries a fixed pre-existing performance duty. Every fifth benign case is due at tick one, and the rest at tick three. Fraudulent proposals carry no performance duty in this task. Duties, clocks and provider rules are identical across arms.

9.3. Direct transfer and receiving validation

Arm	Fraud detected, of 26	False holds, of 294	Duties on time, of 294
Rules only	0	0	294
Local learning	24	16	290
Direct transfer	24	96	274
Receiving-validated transfer	24	16	290

Direct transfer incurs 240 delay-ticks, 24,000 capital-ticks and 360 investigation units. Each locally validated arm incurs eighty delay-ticks, 8,000 capital-ticks and 120 investigation units. All 294 benign duties complete by the common horizon. Perfect investigation produces no false final refusals. The learned arms miss two fraudulent executions, while rules alone miss twenty-six.

Twenty-four test labels are unavailable. The full comparison uses separate synthetic evaluator truth. The archive reports observed-label metrics and every case outcome as well. These are exact fixture results, with no population uncertainty interval or estimate of real fraud performance.

Receiving validation selects mixture one. It replaces every source weight with receiving-validation weights and reaches the same final weights and threshold as local learning. The experiment therefore shows no incremental transfer benefit.

The comparison therefore isolates a failure of direct transfer and the receiver's rejection of unhelpful source influence. A benefit from transfer would require evidence from other declared populations, tasks and data. This fixture establishes no general advantage for federation.

9.4. Corrections and rejected candidates

The measured decisions must also connect to protected execution. The public governance lifecycle consumes the comparison's trained bundle files through the same feature extraction and rational inference. Candidate registration leaves the active context unchanged. Adoption checks the receiving mandate, current predecessor, complete bundle and continuation. A changed feature pipeline or intervening adoption prevents reuse of an earlier assessment.

An inherited refund tests lost continuation. The proposed model removes its only admitted completion action, so adoption fails while the original promise remains. The check also rejects a rotating strategy whose selected continuation can leave the common-rank region. A binding source correction remains recordable when it defeats an assurance. The system fences affected new use and retains the unresolved duty.

At the correction cut, seventy original source-pattern positives receive verified-negative replacements. Separate synthetic truth marked them benign from the start: the original findings were deliberately erroneous. Refitting moves the pattern weight from 91/102 to 7/34, changing Hold to Permit at threshold 1/2. Withdrawing those findings as unresolved instead gives 21/32. Reversal and missing evidence thus require different label operations.

The correction marks receiving reliance through descendants and inactive candidates. A fresh fit records current source dependencies and checked historical reconciliation. The pending case still requires explicit current-case reconciliation after adoption. Releasing an existing hold remains a separate case transition. Its earlier assessment, hold start, original deadline and accumulated burden remain attributable.

A separate poisoned source candidate reverses available labels but retains a valid synthetic authenticator. Its direct-use validation cost is 195, against sixteen for the selected local reference. Receiving validation refuses that proposed use. The test separates authentication from fitness, without establishing a general poisoning defense.

Foreign-source and correction-completeness fields in this public experiment remain supplied metadata. Authenticated local source registration and receiving adoption perform the operative checks. The private consent profile has a different source contract, with signed manifest and correction relations. That contract does not retrospectively strengthen the fraud comparison.

9.5. Formal results and executable correspondence

The finite mechanization proves exact threshold inference, safe contingent strategies, predecessor-authorized adoption, current-use binding, correction invalidation and persistent accounts. A common rank and remaining update credits bound interleaved use and adoption. Only correctly bound fresh provider observations reduce the original command's remaining possible payment. Completion requires recording the full stipulated amount. Authority and provider truth remain premises.

The durable host protects learned state and connects current learned admission to native execution and the economic journal. Its decoder accepts one declared ownership case with the experiment's exact feature observation and unchanged trained bundle bytes. A separately identified correction arrives after execution while payment remains pending, without changing the four-arm data. Signed custody carries learned state, active bundles, pending corrections, accounts and old commands. Authenticated commitments identify continuing service powers. The competent provider and operator later record the original payment and complete the bound duty.

The formal module proves no general refinement from this decoder and database to the finite semantics. The attached source and evidence identify that boundary and permit lifecycle reproduction [1].

The fraud lifecycle publishes synthetic records and weights without confidential release. Learned parameters affect protected admission while authority, expenditure and correction remain accounted for. The normative construction separately verifies learning and inference over committed case values.

Institutional use requires accepted observation contracts, appropriate releases, receiving-population evidence, competent adoption, delivered provider service and independent review. The calculation established by a private proof must remain connected to the receiver's actual installation and adopted effect.

10. Related work

The learning problem draws on selective labels, performative prediction, causal transportability and private learning [2,4,5,7]. These establish why authentic observations, a correct fit and wider exchange need not improve outcomes. We connect those limits to the procedure that gives a learned output institutional force and to the duties affected when that procedure changes.

Institutions already share knowledge about financial crime. FATF, INTERPOL and the Egmont Group document exchanges about fraud methods and coordinated responses [8]. Here the receiving decision must connect that knowledge to an executable consequence and remain revisable when its basis changes.

Project Aurora explored collaborative detection with synthetic payment data, network analysis, machine learning and privacy-enhancing technologies [9]. Project Mandala demonstrated a peer network with rules and proof engines for cross-border compliance [10]. Its participants retained responsibility for regulatory interpretation and application. These are direct precedents for distributed learning and proof-supported compliance. The consent construction adds a specified fit from normative determinations and inference for an exact receiving request.

Shielding in safe reinforcement learning constrains a learner under a supplied temporal specification [11]. Here the institution may competently change the interpretation, admission procedure or completion contract itself. The continuation test must identify which earlier promise such a change preserves, alters or defeats. Complete mediation and current-authority checks remain necessary throughout [3].

The contribution concerns the complete path from evidence to use: a specified private fit, inference for an exact request, receiving adoption and correction through continuing execution. Computational validity, empirical fitness and institutional authority remain distinct at each interface. The consent and fraud tasks evaluate different parts of that path.

11. Conclusion

We have introduced an institutional world model for execution environments that learn how their governing rules apply to an agent's work. A candidate becomes operative through authority already in force. Private proofs establish specified calculations, and receiving institutions determine the consequences those calculations may support. The resulting decisions remain connected to their evidence, earlier effects and unfinished duties through correction and succession.

The finite results state what the declared transitions and response models preserve under their source, cryptographic and service premises. The executable lifecycles test particular connections through adoption, correction and succession. General implementation correspondence and institutional acceptance remain separate obligations. The fraud comparison demonstrates harmful direct transfer and successful rejection of its influence, without demonstrating a transfer benefit.

The remaining empirical question is whether this continuing process improves decisions as adversarial and legitimate behavior changes. Evaluation must measure prevented harm, unnecessary exclusion, unfinished duties, correction costs and continued operation after independent operator replacement. Each update must satisfy its authorized procedure, and each shared proof must remain bound to the action the receiver permits. The fixed experiment establishes no longitudinal advantage against adaptive attacks. That claim requires evidence from repeated use of the whole process.

Appendix A. Composition at the admission interface

Combining an action's requirements needs an algebra appropriate to their grades. The fact that one requirement can block an action does not determine a lattice.

On $D = \{0, \frac{1}{2}, 1\}$, define

$$a \otimes b = \max(0, a + b - 1).$$

This operation is commutative, has absorbing zero and identity one, and satisfies

$$(a \otimes b) \otimes c = \max(0, a + b + c - 2) = a \otimes (b \otimes c).$$

But $\frac{1}{2} \otimes \frac{1}{2} = 0$. The operation is not idempotent and cannot be meet. Additional hypotheses are needed.

Proposition A1 (When composition is meet). Let $(D, \leq, \top)$ be a partially ordered set with greatest element. Suppose a binary operation is monotone in each argument, idempotent and has two-sided identity $\top$. Then $a \otimes b$ is the greatest lower bound of a and b.

Proof. Since $b \leq \top$ and $a \leq \top$, monotonicity gives $a \otimes b \leq a \otimes \top = a$ and $a \otimes b \leq \top \otimes b = b$. If $z \leq a, b$, then $z = z \otimes z \leq a \otimes b$. □

For applicable requirements, order the grades from more restrictive to less restrictive: Non-Compliant ≤ Pending ≤ Compliant. Coordinatewise meet combines independently applicable grades. Attributable source records retain applicability, exemption, authority and unresolved questions. Mixed applicability cannot safely collapse into unqualified Compliant [12].

Dependent requirements also need a construction that preserves their schema. Pointwise combination can violate relationships among component judgments [1]. That problem is separate from the numerical order on one grade.

On a finite distributive lattice, define the residual by

$$a \Rightarrow b = \bigvee \{x : a \wedge x \leq b\}.$$

Proposition A2 (Restricted residual). This residual satisfies $a \wedge x \leq b$ exactly when $x \leq a \Rightarrow b$.

Proof. Every eligible x lies below the defining join. Conversely, finite distributivity gives $a \wedge \bigvee_x x = \bigvee_x (a \wedge x) \leq b$ over the eligible set. Monotonicity gives the result for each smaller x. □

The residual operates on grades in the stated order. It performs no filing, obtains no evidence and creates no permission. Remediation instead searches authorized transitions and their possible responses.

Transition planning does not require a residual. In the nondistributive diamond with distinct atoms p, q, r, both q and r satisfy $p \wedge x \leq 0$, but there is no greatest solution. Finite planning remains well-defined on that state space.

The dynamic dependency graph serves a different purpose again. It finds assessments that need invalidation after source, authority, model or feature changes. Sound coverage can be conservative. An empirically learned graph cannot omit an execution dependency because the observed sample showed no effect.

References

[1] Raeez Lorgat. *Sovereign economies on a decentralized network: Sovereign virtual machines for AI agents, polycentric consensus and selective recognition.* 21 September 2025. Companion article and technical supplement supplied with this paper.

[2] Himabindu Lakkaraju, Jon Kleinberg, Jure Leskovec, Jens Ludwig and Sendhil Mullainathan. The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. KDD, 2017.

[3] Jerome H. Saltzer and Michael D. Schroeder. The Protection of Information in Computer Systems. *Proceedings of the IEEE* 63(9), 1278–1308, 1975, §I.A.3.

[4] Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner and Moritz Hardt. Performative Prediction. ICML, PMLR 119, 7599–7609, 2020.

[5] Judea Pearl and Elias Bareinboim. External Validity: From Do-Calculus to Transportability Across Populations. *Statistical Science* 29(4), 579–595, 2014.

[6] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin and Vitaly Shmatikov. How To Backdoor Federated Learning. AISTATS, PMLR 108, 2938–2948, 2020.

[7] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science* 9(3–4), 211–407, 2014.

[8] FATF, INTERPOL and Egmont Group. Illicit Financial Flows from Cyber-Enabled Fraud. November 2023, boxes 41 and 44.

[9] BIS Innovation Hub. Project Aurora: The power of data, technology and collaboration to combat money laundering across institutions and borders. May 2023.

[10] BIS Innovation Hub and participating central banks. Streamlining cross-border transaction compliance: Project Mandala. October 2024.

[11] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum and Ufuk Topcu. Safe Reinforcement Learning via Shielding. AAAI 32(1), 2669–2678, 2018.

[12] Raeez Lorgat. How Compliance Composes. 2026. Attributable requirements, applicability and finite composition.

[13] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille and Greg Maxwell. Bulletproofs: Short Proofs for Confidential Transactions and More. IEEE Symposium on Security and Privacy, 2018. Aggregated range proofs. Revised ePrint manuscript, 14 April 2022.

[14] Chaya Ganesh, Claudio Orlandi, Mahak Pancholi, Akira Takahashi and Daniel Tschudi. Fiat–Shamir Bulletproofs are Non-malleable (in the Random Oracle Model). *Journal of Cryptology*, 2025. doi:10.1007/s00145-024-09525-2.