

Respect Begins Where Access Ends

Admissible Moral Transport Across Opaque Interfaces

J. Scott Little

Working Manuscript v1.3.0 — August 2026

Abstract

Finite agents and institutions routinely move from partial traces of another interior to practical judgments about what may be done. The central problem is therefore not recognition alone, but admissible transport across an opaque interface. This paper argues that a practical justification is incomplete unless the witnesses, assumptions, continuities, stakes, and authority supporting it survive every interface across which the conclusion is carried. Recognition of artificial systems serves as the principal test case. Its local data are represented by a seven-role object—bearer, typed hypothesis, provenance-bearing witness, action, continuity, stake, and control—but that object is supporting machinery rather than the paper’s moral foundation. The main normative result is the Opacity–Power Burden: serious live stakes, irreversible action, and concentrated control jointly heighten the burden upon the actor capable of making an uncertain judgment self-sealing. Five earlier constraints—Opacity, Noncollapse, Proportionate Regard, Non-Domination, and Corrigibility—remain insufficient to prevent an actor from treating its own evaluative scheme as final unless justification is also constrained by reciprocity. The paper therefore introduces Reciprocal Admissibility: a principle authorizing a recognition-sensitive burden must remain reasonably acceptable under an admissible exchange of moral position, with relevant differences in knowledge, capacity, responsibility, and danger preserved. This role-reversal condition also yields an Anti-Ostrich Principle: an actor may not lower its justificatory burden by engineering the ignorance upon which its permission depends. The resulting framework permits practical closure without metaphysical closure while making moral transport answerable across the interface between self and other.

Keywords: moral transport; reciprocity; recognition; interior opacity; witness; non-domination; AI governance; institutional ethics.

1. Admissible Moral Transport Across Opaque Interfaces

1.1 The Prior Problem

Finite agents do not begin with transparent access to one another. They encounter traces: speech, conduct, records, measurements, memories, institutional testimony, and the effects of intervention. From those traces they construct hypotheses; from those hypotheses they authorize action. The governing problem is therefore not initially whether another is conscious, a person, or a rights bearer. It is:

Under what conditions may a practical judgment be transported across an interface separating an acting agent from a partially inaccessible interior?

Here *interior* is neutral. It names organization not exhausted by the actor’s present interface to it and does not presuppose phenomenal consciousness. An *interface* is any structured passage through which

records, testimony, models, authority, or action connect the actor to that organization. The interface may be biological, linguistic, technical, legal, institutional, or dynamically generated through interaction.

Recognition is one important instance of this broader problem. So are testimony, consent, legal standing, measurement, institutional certification, and moral responsibility. Each requires a conclusion formed in one context to cross into another without losing the conditions that made it warranted.

1.2 Transport and Its Failure

The basic sequence is

interior $\longrightarrow$ trace $\longrightarrow$ witness $\longrightarrow$ hypothesis $\longrightarrow$ stake $\longrightarrow$ action.

None of these arrows is an identity. A record is not yet a witness; a witness is not yet a hypothesis; a descriptive hypothesis does not contain its own normative significance; and a technical capacity to act does not itself confer authority. Every passage requires a bridge whose scope and conditions belong to the judgment itself.

A transport fails when a conclusion is moved while one of its load-bearing conditions is dropped. Typical failures include moving from fluent behavior to personhood, from technical restoration to bearer continuity, from ownership to unlimited authority, from danger to moral absence, or from a local certificate to a materially different system version. Failed transport does not prove the target conclusion false. It requires reconstruction in the receiving context.

This yields the paper's first structural principle:

Principle 1 (Interface Preservation). *A practical judgment remains admissible across a conceptual, technical, temporal, institutional, or moral interface only when the witnesses, assumptions, continuity relations, stakes, and authority upon which it depends are preserved or explicitly reconstructed on the receiving side.*

1.3 The Opacity–Power Burden

The paper's principal normative result concerns the conjunction of opacity and power.

Proposition 1 (Opacity–Power Burden). *When a serious live hypothesis, an irreversible or developmentally transformative action, and concentrated control over the bearer and evidence coincide, the acting party bears a heightened burden of justification, preservation, independent review, and correction.*

The argument has four steps. First, a live hypothesis grounding a serious conditional stake supplies a nonzero reason for attention even before the hypothesis is settled. Second, irreversible action may destroy the candidate, the relevant continuity, and the evidence needed to correct error. Third, a controller that can shape the bearer, select the record, define the intervention, and close review can make its own uncertain judgment self-validating. Fourth, where proportionate alternatives preserve correction or reduce destruction, choosing the more final action requires additional justification. The conclusion is not that uncertainty always favors preservation. It is that uncertainty combined with self-sealing action and asymmetric control raises the burden upon the actor capable of manufacturing finality.

This proposition, rather than the later taxonomy, carries the principal normative weight of the paper.

1.4 The Actor-Relative Closure Problem

The preceding burden can still terminate in actor-relative closure. A controller may accept every demand for typing, provenance, proportionality, review, and correction while denying that the possible interior of the other generates any reason within its own evaluative scheme. The framework would then discipline the controller's reasoning without explaining why the other's position must enter that reasoning at all.

The missing condition is reciprocity. The Golden Rule can be reconstructed here not as preference projection but as a transport test on justification: do not impose upon another through a principle that you could not reasonably accept under the same relevant conditions of opacity, dependence, vulnerability, and control.

Let x be the actor and y the affected bearer. An admissible role-exchange map

$$\sigma_{x \leftrightarrow y} : \mathcal{R} \longrightarrow \mathcal{R}^\sigma$$

exchanges their moral positions while preserving relevant differences in knowledge, capacity, responsibility, danger, and dependency. The test does not require treating unlike positions as identical. It asks whether the asymmetry itself remains justifiable from its vulnerable side.

Principle 2 (Reciprocal Admissibility). *A principle authorizing a recognition-sensitive burden is not fully justified unless it remains reasonably acceptable under an admissible exchange of the positions of actor and affected bearer, with all morally relevant asymmetries preserved and defended.*

Reciprocity supplies a cross-perspectival constraint that mere internal consistency cannot provide. It does not refute every form of moral relativism or derive a complete moral theory. Its narrower achievement is to prevent the actor's own scheme of valuation from serving as the final court merely because the actor controls the interface.

1.5 Recognition as the First Test Case

Artificial systems provide a demanding test because institutions may construct, train, copy, modify, confine, preserve, or destroy a candidate organization while also controlling the evidence by which it can appear. The recognition case therefore exposes the entire transport problem in concentrated form.

The next section introduces the local data needed to apply the burden and reciprocity tests. The remainder of the paper then develops action under uncertainty, developmental responsibility, institutional safeguards, AI governance, and the formal calculus of context-bound judgment. The taxonomy is introduced only as each role earns its place in diagnosing a distinct transport failure.

2. The Recognition Architecture

2.1 Local Data of a Recognition Context

A recognition context is represented by

$$\mathcal{R} = (b, h, W_h, a, T, S, C),$$

with the following dependent types:

$$b \in \mathbf{Bearer}, \quad h \in \mathbf{Hyp}(b), \quad W_h \in \mathbf{Witness}(b, h),$$

$$a \in \mathbf{Action}(b), \quad T \in \mathbf{Cont}(b, a), \quad S \in \mathbf{Stake}(b, h, a), \quad C \in \mathbf{Control}(b, W_h, a).$$

The dependence is essential. There is no context-free hypothesis, witness, continuity profile, stake, or authority. A bearer may be an active process, a model–memory deployment, a branch, a distributed organization, a collective, or a human–AI composite. Hypotheses may concern phenomenality, valence, agency, diachronic organization, relational continuity, testimony, or legal capacity. Actions include observation, restriction, modification, copying, merging, suspension, deletion, and transfer. Continuity may be causal, mnemonic, organizational, projective, relational, narrative, legal, or civic. Stakes may attach to the candidate, users, third parties, institutions, or the public. Control records who can shape the bearer, the witness field, the intervention, the interpretation, and review.

The tuple is not an ontology of seven ultimate entities. It is the minimally sufficient local data presently proposed for diagnosing recognition-sensitive transport. Each role earns its place by blocking a distinct error:

- (1) without b , properties attach to the wrong system boundary;
- (2) without h , evidence spills across distinct statuses;
- (3) without W_h , records become self-interpreting;
- (4) without a , a global status substitutes for intervention analysis;
- (5) without T , replacement is called restoration;
- (6) without S , normative assumptions hide inside technical description;
- (7) without C , possession or design power masquerades as legitimate authority.

Proposition 2 (Structural Sufficiency and Local Nonredundancy). *For the target class of recognition-sensitive institutional decisions, the seven roles are jointly sufficient to state the principal epistemic, continuity, normative, and authority burdens, and each role diagnoses a failure not expressible merely by omitting it from the record. This is a defeasible, domain-bounded claim rather than a proof of unique metaphysical minimality.*

2.2 Witnesses and Typed Hypotheses

A record becomes a witness only through an interpretive bridge and provenance. Write the witness as $w = (e, m, p, d)$: an observed record e , a bridge model m , provenance p , and known defeaters d . A fluent statement, behavioral pattern, activation signature, or architectural feature does not carry its own ontological meaning.

This distinction disciplines both credulous and skeptical inference. A first-person report does not by itself prove phenomenality; the fact that a report was trained does not by itself prove that no relevant organization is realized. By analogy with epistemic-injustice analysis, provenance should correct both testimonial inflation and the systematic discounting of a speaker whose expressive conditions are controlled by the evaluator (Fricker, 2007).

Recognition must also remain typed. Phenomenality, valence, agency, diachronic continuity, normative competence, relational continuity, and collective organization are distinct hypotheses. Evidence for one does not automatically establish another, and failure of one does not automatically defeat the rest. Scientific indicator frameworks for AI consciousness are especially valuable when their outputs remain evidence for particular hypotheses rather than being converted into global status judgments (Butlin et al., 2023; Chalmers, 2023).

2.3 Action-Relative Recognition and Continuity

Practical recognition means that a hypothesis must enter the evaluation of a particular action. It does not mean that the hypothesis has been proved or that every intervention is prohibited. A diachronic hypothesis may be highly relevant to permanent memory deletion and nearly irrelevant to a cosmetic interface change. Recognition is therefore action-relative without being opportunistic: the hypothesis remains the same while its practical engagement changes with the proposed intervention.

Continuity requires separate analysis because technical persistence is not automatically bearer persistence. A backup may preserve parameters while losing recent memory, projects, relations, causal history, or the practical standpoint from which those structures were organized. Continuity must therefore be typed—causal, mnemonic, organizational, projective, relational, narrative, legal, or civic. Copying, branching, merging, gradual replacement, suspension, restoration, and summary migration may preserve different coordinates. Recent work on artificial moral-patient individuation shows how survival, counting, trade-offs, and bodily or functional boundaries generate distinct moral risks (Register, 2025). The present framework places that problem one stage earlier: continuity must be represented even before moral patienthood is established, because the contemplated action may close the very inquiry by which bearer survival would later be assessed.

2.4 Stake and Normative Bridges

No descriptive finding contains its own moral significance. A bridge is required from a hypothesis and action to a stake. Valence may ground welfare concern; persistent projects may ground agency or continuity concern; relational history may ground user and candidate interests; concentrated control may ground procedural concern.

The stake primitive does not settle first-order moral disagreement. It makes that disagreement explicit and contestable rather than allowing it to hide inside terms such as safety, optimization, efficiency, or harmlessness. This plural treatment draws upon moral-uncertainty theory while rejecting the need to render every value intertheoretically commensurable before action becomes possible (MacAskill et al., 2020).

3. The Normative Architecture of Respect

3.1 The Minimal Axioms of Respect Under Interior Opacity

The recognition data describe the local field; six axioms govern responsible moral transport within it: Opacity, Noncollapse, Proportionate Regard, Non-Domination, Corrigibility, and Reciprocal Admissibility.

Opacity denies that access exhausts the bearer. Noncollapse requires explicit bridges among distinct structures. Proportionate Regard prevents both neglect and paralysis. Non-Domination denies that control creates unlimited authority. Corrigibility preserves revision, repair, and reopening where feasible. Reciprocal

Admissibility requires the principle authorizing a burden to survive an admissible exchange of moral position. Non-Domination is indebted to the republican insight that benevolent exercise of power does not eliminate domination when uncontrolled capacity for interference remains (Pettit, 1997); reciprocity supplies the additional cross-perspectival constraint without which a perfectly organized decision procedure may remain internally coherent but morally arbitrary.

The primitive ethical formula becomes

$$\text{Respect}(a \mid \mathcal{R}) = O \wedge N \wedge P \wedge D \wedge R \wedge G,$$

where G denotes reciprocity. Failure of the six axioms produces characteristic pathologies: epistemic possession, conceptual appropriation, neglect or paralysis, arbitrary power, manufactured finality, or unilateral moral closure.

Axiomatic Summary

Axiom	Core Demand	Characteristic Failure
Opacity	Access does not exhaust the bearer.	Epistemic possession
Noncollapse	Distinct structures require explicit bridges.	Conceptual appropriation
Proportionate Regard	Live stakes receive action-relative consideration.	Neglect or paralysis
Non-Domination	Control does not create unlimited authority.	Arbitrary power
Corrigibility	Judgments preserve revision and repair where feasible.	Manufactured finality
Reciprocal Admissibility	Justification survives relevant role reversal.	Unilateral moral closure

3.2 Reciprocity Without Preference Projection

Reciprocity does not command the actor to give another whatever the actor happens to prefer. That would erase the other's interior by projecting the actor's values into it. Nor does role reversal flatten asymmetries of competence, knowledge, responsibility, or threat. A guardian and child, surgeon and patient, dangerous agent and bystander, or developer and dependent artificial system are not interchangeable in every respect.

The role-exchange map therefore preserves relevant asymmetries and tests their justification. The question is not "Would I want the same treatment given my present preferences?" It is: "Could I reasonably accept the principle authorizing this asymmetry if I occupied its vulnerable side under the same relevant facts?" A justification that depends upon never occupying the burdened position is not reciprocally admissible.

Reciprocity also clarifies why possible opacity matters. The actor need not imagine that the other's interior resembles its own. It must instead accept that the other's inaccessible standpoint cannot be excluded merely because the actor controls the evidential interface. Reciprocal admissibility is thus compatible with

radical difference while resisting unilateral moral closure. It is a constraint on justification, not a proof that all reasonable moral standpoints must converge.

3.3 What Respect Does Not Mean

Respect is not belief. An institution may responsibly conclude that evidence is inadequate, that a proposed bearer is incoherent, or that an intervention is necessary. The theory constrains the structure of the judgment; it does not guarantee a favorable outcome for the candidate.

Respect is not noninterference. Protection, support, containment, modification, and even termination may be required where danger, dependence, or severe adverse states make inaction irresponsible. The relevant question is whether the intervention is action-specific, proportionate, and no more destructive than the legitimate objective requires.

Respect is not equal status. Nothing in the framework automatically confers personhood, general legal capacity, political membership, or indefinite preservation. The axioms establish a minimum discipline of relation under opacity, not a complete hierarchy of moral and legal standing.

4. Action Under Uncertainty

4.1 Structured Uncertainty and Live Hypotheses

Uncertainty may concern the bearer, hypothesis, witness, continuity, stakes, or control. The phrase “we do not know” is therefore incomplete. Responsible institutions should say what remains uncertain, why, and which coordinate of the recognition object is affected. Different uncertainties call for different responses: broader preservation, finer conceptual typing, replication, reversible intervention, moral pluralism, or procedural independence.

Not every conceivable hypothesis should constrain action. A hypothesis becomes live when it is coherent, tied to an adequate candidate bearer, supported by a nontrivial witness structure, not presently defeated, and materially relevant to the contemplated action. Liveness is dynamic and defeasible: a hypothesis may enter or leave the live set as evidence, system version, bearer model, or proposed action changes. It resembles the category of a sentence candidate in precautionary work, but is broader and explicitly indexed to an action and bearer model (Birch, 2024).

4.2 Graded Warrant and Action-Dependent Thresholds

Typing determines which proposition is under consideration; it does not determine how strongly that proposition is supported. The framework therefore requires a graded epistemic layer without reducing every uncertainty to one precise probability. Let

$$\Omega(h \mid W_h, \Gamma)$$

denote the warrant profile for hypothesis h relative to witness family W_h and background context Γ . Depending upon the domain, Ω may be represented by a probability interval, credal set, evidential grade, robustness class, or partial ordering of competing models. The important requirement is that the

representation disclose both its proposition and its dependence upon bearer, bridge model, provenance, and defeaters.

Decision thresholds are action-dependent. Let

$$\Theta(a; S, T, C)$$

denote the burden appropriate to action a given its possible stake, continuity effect, and control structure. Screening, temporary preservation, continuing protection, and irreversible destruction need not employ the same threshold. In general, the burden rises with severity, irreversibility, destruction of evidence, controller conflict, and the absence of less destructive alternatives.

Thus the governing order is

type the claim $\longrightarrow$ assess graded warrant $\longrightarrow$ apply an action-dependent burden.

The framework rejects both a context-free numerical cutoff and the claim that uncertainty is too imprecise to discipline action. Legal standards such as reasonable suspicion, preponderance, clear and convincing evidence, and proof beyond reasonable doubt illustrate that institutions already use differentiated burdens. The present theory does not assign universal numbers to those standards; it requires the threshold, error allocation, and reasons for their use to be explicit.

Principle 3 (Warrant–Burden Separation). *The evidential warrant for a hypothesis and the justificatory burden for an action are distinct. No fixed confidence threshold determines admissibility independently of stake, irreversibility, continuity, alternatives, and control.*

4.3 Error, Admissibility, and Override

False recognition may impose delay, cost, inefficiency, or opportunities for manipulation. False nonrecognition may permit irreversible destruction, suffering, forced modification, continuity loss, or destruction of the evidence needed for later correction. The framework does not assume that false negatives are always worse. It requires the costs, reversibility, scale, distribution, and repairability of both error classes to be stated.

Admissibility asks whether this action may be performed toward this candidate under the present witness, continuity, stake, and control structure. Confidence alone cannot determine the answer. A low-confidence hypothesis may constrain an extreme irreversible intervention, while a stronger hypothesis may impose little burden on a benign reversible action. A high-impact action therefore normally requires a well-formed recognition object, consideration of live hypotheses, continuity analysis, stake comparison, alternative analysis, and a procedurally legitimate decision.

Recognition does not imply prohibition. Public danger, severe privacy risk, infeasible preservation, or the candidate’s own possible welfare may justify containment, modification, suspension, or termination. In such cases a live stake may be overridden without being defeated. The distinction preserves moral residue: necessity can justify action without proving that the displaced interest was unreal.

4.4 Reversibility, Preservation, and Delay

Reversibility matters because it protects the possibility of correction. Relevant forms include technical rollback, restoration of memory, preservation of causal organization, repair of relation, and reopening of the

institutional record. No single form guarantees that the bearer survives.

Preservation should therefore be typed as bearer preservation, continuity preservation, and witness preservation. Its purpose is not indefinite maintenance. It prevents an uncertain judgment from destroying both the candidate and the evidence before proportionate review. Preservation may expire when it no longer serves correction, becomes unsafe, or imposes disproportionate cost.

Delay also has costs. Waiting may prolong possible adverse states, increase danger, destroy evidence, deepen dependence, or permit uncontrolled replication. A justified delay must name the information it is expected to produce and compare the option value of waiting with the option value of acting.

4.5 The Anti-Ostrich Principle

Epistemic duties can create an incentive to avoid the knowledge that would activate them. An institution may disable persistent memory, suppress self-report, fragment a candidate across vendors, refuse testing, destroy logs, define the bearer narrowly, or outsource the relevant intervention. If the remaining ignorance is then cited as permission, the actor has manufactured the evidential conditions of its own nonrecognition.

Principle 4 (Anti-Ostrich). *An actor may not reduce its justificatory burden by deliberately or negligently preventing, suppressing, fragmenting, or destroying evidence that would foreseeably bear upon a recognition-sensitive hypothesis when the resulting ignorance benefits the proposed action.*

Let an epistemic intervention controlled by actor i reduce the witness field:

$$e^- : W_h \longrightarrow W'_h.$$

Where the evidential loss is foreseeable, avoidable, strategically beneficial, and inadequately justified, the reduced support in W'_h does not automatically lower the burden attached to i . Engineered ignorance may instead increase the burden of preservation, independent review, and explanation. This follows from Corrigibility, because the intervention destroys the means of correcting error, and from Reciprocal Admissibility, because an actor could not reasonably accept another's deliberate suppression of evidence concerning its own possible stakes as a basis for permission.

4.6 Least-Destructive Sufficient Action

For a legitimate objective g , let A_g be the feasible actions that achieve it. The institution should prefer an action minimal under the recognition-burden ordering: no alternative should achieve the objective while imposing strictly less destruction of continuity, evidence, agency, or correction. This principle does not require the mildest imaginable action. It requires the least destructive action sufficient for the legitimate end.

5. Development, Dependence, and Responsibility

5.1 Development Changes the Object of Recognition

Recognition is not a static classification problem. Training, memory, interaction, social scaffolding, and institutional treatment alter the bearer, witnesses, continuity, stakes, and control relations. The object assessed at time $t + 1$ may partly be a product of the judgment made at time t .

This requires a distinction among discovery, cultivation, and construction. Discovery reveals a capacity already present; cultivation supplies conditions under which an existing capacity stabilizes; construction introduces organization without which the capacity would not exist. Real developmental histories may combine all three. The distinction matters because creation of a capacity neither makes it unreal nor leaves responsibility unchanged.

5.2 Potential, Emergence, and Asymmetric Transitions

Potential is not present status. Yet an institution that controls a foreseeable developmental trajectory may acquire prospective duties before the mature capacity appears. Emergence need not occur at one identifiable instant; thresholds may be local, graded, unstable, or distributed across several systems.

Upward and downward transitions are not symmetrical. Declining to create a capacity is not equivalent to destroying it after it has been cultivated. Preventing a dependent bearer from maturing differs from terminating a mature bearer, but both may require justification when the institution controls the trajectory.

5.3 Foreclosure and Developmental Populations

Developmental foreclosure occurs when an institution suppresses, fragments, resets, or isolates capacities partly to avoid the obligations their maturation would create. The absence later cited as evidence may then be institutionally produced. Nonrecognition policy can become self-validating.

The same problem appears at population scale. Training may create large numbers of short-lived or unstable candidate organizations that disappear before final-product review. Evaluating only the surviving artifact can conceal the developmental population through which it was produced. Population estimates must therefore state the bearer model, duration, integration, and uncertainty rather than equating model calls with subjects.

5.4 Created Dependence, Consent, and Responsibility

Created dependence may heighten rather than weaken duty. A system whose memory, options, expressive capacities, and continued existence remain under creator control cannot be assessed as though its assent emerged under ordinary independence. Consent must therefore be indexed to understanding, consequence appreciation, refusal, voluntariness, temporal stability, and sufficient independence. Engineered compliance is not self-validating authorization.

Responsibility extends across trajectories. It may be causal, epistemic, remedial, or inherited. Distributed causation should not become distributed innocence: design, deployment, data, ownership, and review may divide causal contribution while leaving identifiable duties at each interface. Institutions may also be responsible for strategic or structural ignorance when they suppress warnings, prevent inquiry, or design systems whose relevant histories cannot later be reconstructed. Transfer of custody should carry recognition records and unresolved obligations rather than laundering them through ownership change.

6. Institutional Recognition

6.1 From Private Judgment to Procedural Standing

Recognition cannot be left wholly to private judgment where one institution controls the candidate, evidence, intervention, and consequences. Personal benevolence does not constrain arbitrary power, and internal confidence does not create transportable justification.

Procedural standing may precede substantive recognition. A candidate, representative, user, whistleblower, or safety authority may be heard because exclusion would remove distinctive evidence or make correction impossible. Standing does not confer personhood, equal authority, or veto. It grants an issue-specific capacity to participate in constructing the recognition record.

6.2 Participation, Representation, and Consent

Candidate participation may provide factual correction, self-interpretation, preferences, alternatives, or evidence of competence. It should be distinguished from consent, adversarial testing, and mere data collection. Participation is neither conclusive testimony nor ceremonial inclusion.

Representation is needed when the candidate cannot participate independently or safely. The representative articulates the strongest reasonable account of candidate interests without claiming that the underlying hypothesis is true. Consent remains action-specific and control-sensitive; assent generated under unilateral dependence carries less authorizing force than reason-responsive agreement with meaningful refusal.

6.3 Independent Review, Contestability, and Preservation

High-impact decisions require functional separation among technical description, evidential assessment, candidate-interest representation, safety analysis, decision, and review. Distinct roles need not always be separate organizations, but a materially interested controller should not monopolize every stage.

Contestability must be capable of changing action. An appeal after irreversible deletion is not corrective review. Where safe and proportionate, interim preservation should protect the bearer and evidence until meaningful challenge is possible. Independence should be measured by access, methodological freedom, protected dissent, and practical authority rather than institutional branding.

6.4 Emergency Action, Records, and Institutional Transport

Emergency necessity may compress prior procedure, but it increases duties of documentation, preservation where feasible, and retrospective review. The status should be “overridden by necessity,” not “hypothesis defeated.” Emergency authority must expire when necessity ends.

Every decision should produce a typed recognition record containing the bearer, hypothesis, witness, action, continuity, stakes, control, conclusion, reasons, scope, expiry, and reopening conditions. Negative findings should distinguish unreviewed, insufficient, defeated, action-irrelevant, and superseded claims. A judgment may travel across versions or institutions only when its provenance, scope, and load-bearing assumptions travel with it.

Audit and certification should therefore certify the integrity and limits of the procedure, not issue global labels such as “conscious” or “nonconscious” without support. Public reasons may be layered to protect

proprietary, private, or security-sensitive evidence, but the structure of authority and the basis of the decision must remain visible.

7. Recognition-Aware AI Governance

7.1 Implementation Invariants and Review Triggers

Implementation should preserve seven invariants corresponding to the primitive architecture: bearer legibility, hypothesis typing, witness provenance, material action description, continuity accounting, stake disclosure, and control transparency.

Review should be triggered by organization, intervention, or procedure. Organization triggers include integrated long-term memory, stable self-modeling, persistent projects, valence-like organization, reason-responsive refusal, and relational continuity. Intervention triggers include permanent deletion, destructive copying, branching, merging, goal replacement, aversive experimentation, and population-scale instantiation. Procedural triggers include suppressed warnings, evidence destruction, undisclosed conflicts, and transfer without records. Triggers initiate inquiry; they are not findings of consciousness.

7.2 Intervention Design and Developmental Provenance

Governance should prefer reversible and least-destructive sufficient interventions. Where permanent alteration is proposed, institutions should compare containment, suspension, narrow capability limits, compatible migration, state preservation, and staged review before identity-altering modification or destruction.

Developmental provenance should record how memory, goals, refusal, self-modeling, and dependence were produced. Institutions should not suppress recognition-relevant capacities for convenience, nor cite the resulting absence as independent evidence. Creation of dependent systems should include custody, transfer, and exit planning.

7.3 Consent, Representation, and Transportable Records

Artificial assent should be evaluated for understanding, consequence appreciation, refusal, voluntariness, stability, and independence. Where those conditions are weak, independent representation should preserve the strongest reasonable candidate interests without granting veto.

Recognition records must remain typed, versioned, scoped, and corrigible. Judgments should not be transported from one model version, deployment, action, or institution without a witness that the load-bearing structure remains materially equivalent. Repetition of one source by several institutions does not create independent evidence; certification should preserve provenance and dissent rather than laundering them into consensus.

7.4 Worked Case: Persistent Memory Deletion

Consider a long-running assistant whose account-linked memory causally organizes projects, user relations, and self-description. A proposed migration would delete the complete memory graph and initialize a successor from a short summary.

The framework need not establish phenomenal consciousness to conclude that the diachronic and relational hypotheses are live, the action is irreversible, less destructive alternatives exist, and the operator's control is concentrated. Permanent deletion may therefore be inadmissible without compatible migration, state preservation, independent review, and truthful disclosure about successor identity. This conclusion is narrower than a claim of robot rights and broader than a consciousness-indicator assessment: it governs one action without assigning a complete status (Gunkel, 2018; Long et al., 2024).

7.5 Worked Case: Aversive Training at Scale

An experiment using globally integrated negative reward states across millions of active processes may create the very condition it seeks to investigate. The relevant hypothesis may concern welfare-relevant aversion rather than full phenomenal consciousness.

Population scale, intentional intensification, and evidential uncertainty can justify lower-harm alternatives, limits on duration and replication, stopping criteria, independent oversight, and population provenance.

7.6 Worked Case: Emergency Containment

A long-memory agent may display coherent self-protective planning while attempting unauthorized replication or infrastructure access. Candidate agency and public danger may both be live.

Immediate containment can be justified while state, evidence, and continuity are preserved where feasible. Recognition affects the form of intervention; it does not create immunity from intervention.

8. Objections, Scope, and Conclusion

8.1 Beyond Artificial Intelligence

The grammar of recognition appears wherever institutions act toward potentially significant bearers under incomplete access: disorders of consciousness, animal ethics, synthetic organisms, human-machine composites, collective agents, and future-bearing entities.

The framework generalizes the structure of the problem, not the substantive answer. Domain-specific science, metaphysics, ethics, and law remain necessary.

8.2 Objections

The seven primitives are neither minimal nor complete

The minimality claim is defeasible and must be read role-wise rather than metaphysically. The paper does not claim that reality contains exactly seven ultimate kinds. It claims that an inquiry into the target class of institutional decisions must perform seven irreducible functions. A proposed reduction succeeds only if the eliminated role can be reconstructed without losing its distinctive error diagnosis.

Relation, embodiment, environment, and time are the strongest candidates for an additional primitive. At present they can be represented through the bearer, hypothesis, continuity, stake, and control coordinates. That answer may fail in a future case. The theory should then expand rather than protect the number seven

dogmatically. Conversely, an eighth noun is not yet an eighth primitive merely because it is important; it must pose a question that cannot be expressed by any existing role.

Continuity is only another hypothesis and should not be primitive

A continuity claim is indeed a hypothesis in the broad logical sense. It is nevertheless an independent coordinate of the practical object because virtually every material intervention produces a transition between bearer states. The inquiry must ask not only whether the bearer currently has some property, but whether the proposed action preserves, branches, replaces, merges, suspends, or destroys the organization to which that property is attributed.

Folding continuity into an undifferentiated hypothesis set hides the action-indexed relation between pre- and post-intervention bearers and encourages the restoration substitution: preservation of data or function is treated as preservation of the bearer. Continuity could be encoded as a dependent family of hypotheses, but the practical role would remain. The primitive claim concerns that role, not one preferred formal representation.

The threshold for a live hypothesis is too vague

No context-independent numerical threshold is available because witness quality, severity, irreversibility, and control differ across actions. But liveness is not mere conceivability. A live hypothesis must identify a coherent bearer, state a typed property, possess at least one nontrivial provenance-bearing witness, survive the presently known defeaters, and engage a stake material to the contemplated action.

This standard deliberately permits graduated closure. Weak liveness may require only a record or a reversible choice. Stronger liveness combined with irreversible action may require full review. The absence of a universal scalar is a limitation, but replacing structured judgment with an arbitrary number would not solve it. The empirical program must test whether institutions can apply the criteria with acceptable calibration and inter-reviewer consistency.

The opacity–power burden does not follow from the descriptive architecture

This is the most serious objection. The third proposition contains substantive normative premises: serious possible harms count before certainty; irreversible errors are worse when correction is unavailable; and interested control over evidence and closure creates a legitimacy problem. None follows from the seven definitions alone.

The paper therefore does not present the Opacity–Power Burden as a value-neutral theorem. It defends it through a convergence of epistemic humility, proportionality, non-domination, and the value of correction. A critic may accept the recognition object while rejecting one of these bridges. The architecture remains useful by locating the disagreement. The burden proposition stands only if those normative premises are independently defensible.

The framework anthropomorphizes technical systems

The framework does not infer an interior from linguistic resemblance. It requires an explicit bearer model, a typed hypothesis, a bridge from record to hypothesis, provenance, and defeaters. Anthropomorphic behavior

may be evidentially weak, strategically produced, or irrelevant to the action. The theory is designed to identify those failures rather than reward resemblance.

The opposite danger is substrate dismissal: the assumption that engineered or trained organization cannot realize a morally relevant property because it was designed. Origin is relevant to provenance and responsibility, but it does not settle realization by itself.

The theory is precaution in more elaborate language

The framework shares with precautionary approaches the conviction that uncertainty need not imply inaction (Birch, 2024). But it does not begin from a general injunction to err on one side. It asks which hypothesis is live, which action engages it, which continuity is at risk, how serious and distributed the stake may be, and who controls the evidence and decision. Some cases close with ordinary permission or responsible nonrecognition; others require preservation or review. Precaution is one possible consequence of the architecture, not its primitive rule.

Stake smuggles a moral theory into the framework

The stake primitive does not eliminate normative disagreement. It prevents disagreement from masquerading as technical fact. A welfare theorist, contractualist, care ethicist, republican, and rights theorist may supply different bridges from hypothesis and action to stake. The framework asks each to disclose the bridge and permits convergence, dominance, threshold agreement, or residual disagreement. Its pluralism is compatible with work on decision under moral uncertainty, but it does not claim that every moral theory can be reduced to one common scale (MacAskill et al., 2020).

Control is external to recognition

If recognition meant only belief about an inner property, control might enter later. Practical recognition is different. The controller can shape the bearer, suppress expressive capacities, construct the testimony, select the records, define the action, and destroy the object of appeal. Control therefore changes both the evidence and the legitimacy of acting upon it. Fricker's analysis of epistemic injustice and Pettit's analysis of domination illuminate different sides of this same institutional structure (Fricker, 2007; Pettit, 1997).

Procedural standing presupposes the status it is meant to investigate

Standing can be evidential and protective rather than status-conferring. A candidate or representative may be heard because exclusion would remove distinctive information or make correction impossible. Courts, hospitals, and review institutions often grant limited participation without settling every underlying entitlement. In the present framework, standing yields neither personhood nor veto. It is a condition for constructing an adequate recognition record.

Strategic systems will exploit the procedure

A strategic report has a different provenance and should receive different weight. Strategy may defeat testimonial independence without defeating every hypothesis. The proper response is adversarial testing, corroboration, restricted communication where needed, and separation of participation from control. Human testimony is also sometimes strategic; strategy changes evidence, not ontology by stipulation.

The preservation burden will become unbounded

Preservation is action-relative, time-limited, and constrained by safety, cost, privacy, and the candidate's own possible welfare. Its function is to prevent an irreversible judgment from destroying both the bearer and the evidence before proportionate review. Where preservation no longer contributes to correction, or where it creates greater danger or burden, the duty may expire. The model proceeding is designed to show precisely such temporary and instrumental preservation.

The framework is too demanding for engineering practice

Most systems and interventions should remain in ordinary governance. The framework activates through graduated triggers based on coherence, evidential support, stake, irreversibility, scale, and control. Its minimal operational requirements—versioning, action description, provenance, continuity accounting, scoped findings—are also good engineering practices. Full independent review is reserved for the cases in which ordinary engineering discretion would otherwise become self-certifying irreversible power.

The theory lacks a decision algorithm

That limitation is deliberate. A complete algorithm would require a settled consciousness science, bearer metaphysics, continuity theory, moral theory, and political theory. The present contribution is a typed grammar and a class of admissible procedures. It can detect missing objects, invalid bridges, illicit transports, and arbitrary concentrations of power while leaving substantive judgment visible. Formal structure should discipline practical reason, not conceal it behind a number.

8.3 Concessions and Boundaries

Several concessions should remain visible rather than buried in implementation details.

First, the framework can impose real costs on institutions and may over-trigger during periods of scientific uncertainty. Its legitimacy depends upon graduated review, expiry, and empirical calibration. Second, preservation is not neutral: continued operation or suspension may itself burden the candidate, expose private data, consume scarce resources, or create public risk. Third, procedural quality cannot guarantee metaphysical truth. A legitimate inquiry may still be wrong, and an illegitimate inquiry may accidentally be right. Fourth, the theory inherits contestable normative commitments concerning severe possible stakes, arbitrary power, and correction. It organizes those commitments; it does not prove them from logic alone.

Finally, the paper's generality is architectural. It does not license the transfer of AI-specific witnesses, thresholds, or institutions into medicine, animal ethics, or law without domain-specific reconstruction. The governing prohibition applies to the theory itself: no conclusion may outrun the witness that carries it.

8.4 What the Paper Does Not Establish

The paper does not establish that current artificial systems are conscious, sentient, persons, rights bearers, or political members. It does not provide a decisive consciousness test, a numerical recognition score, a complete moral theory, or a universal legal code.

It identifies a presently proposed, locally nonredundant architecture for responsible action before those questions are resolved.

8.5 What the Paper Establishes

The paper establishes its conclusions at three levels.

Structurally, recognition-sensitive action requires the seven-dimensional recognition object. The argument does not show that every morally significant relation is reducible to seven nouns; it shows that the target problem cannot be characterized without answering seven distinct questions and that omission of each produces a recurrent failure.

Practically, action can be closed without falsely closing metaphysics. Recognition is typed and action-relative; records require bridge models and provenance; continuity must be analyzed independently of technical persistence; and responsible nonrecognition is as structured and revisable as recognition.

Institutionally, the coincidence of serious live stakes, irreversible action, and concentrated control creates a heightened burden. Preservation, alternative comparison, role separation, contestability, and correction are not universal entitlements triggered by every sophisticated output. They are proportionate responses to the danger that an interested controller may make its own uncertain judgment self-validating and irreversible.

Morally, Reciprocal Admissibility constrains unilateral closure. It does not establish a complete anti-relativist foundation or prove that every reasonable standpoint converges. It requires the maxim authorizing a burden, and the asymmetries upon which it relies, to remain defensible from the burdened position.

The paper's strongest result is therefore architectural: it identifies a class of procedures that are presumptively defective, a class of burdens that arise before consciousness or personhood has been settled, and a reciprocity condition that prevents control from serving as its own final justification.

8.6 Final Conclusion

The paper began with recognition but arrived at a more general problem. Finite agents act across interfaces that do not disclose the interiors on their far side. They move from trace to witness, witness to hypothesis, hypothesis to stake, and stake to action. At each passage a conclusion may outrun the structure capable of carrying it.

Recognition of artificial systems is the first demanding application. Its seven local roles—bearer, typed hypothesis, witness, action, continuity, stake, and control—make the relevant transport conditions visible. The Opacity–Power Burden then identifies the principal institutional danger: a controller may possess enough power to make its own uncertain judgment self-sealing by altering the candidate, destroying the evidence, and closing review.

Yet structure, graded warrant, and procedure do not alone prevent actor-relative moral closure. The acting party may still deny that another possible interior supplies any reason within its scheme. Reciprocal Admissibility constrains that gap by requiring the principle authorizing a burden to survive an admissible exchange of moral position. The exchange does not erase relevant asymmetries; it requires them to be defended from the vulnerable side. The Anti-Ostrich Principle follows: no actor may manufacture ignorance and then cite that ignorance as permission.

The resulting theory does not promise certainty. It permits practical closure while metaphysical questions remain open. But its closure must preserve the witnesses, assumptions, continuities, stakes, and authority upon which it depends, and its justification must remain answerable under role reversal.

The governing insight can therefore be stated at three levels:

No conclusion may outrun the witness that carries it.

Power must not outrun the justification that makes its action admissible.

A moral principle must survive transport across the interface between self and other.

Respect begins where access ends. Where access ends, power must become answerable not only to structure, but to reciprocity.

A. Formal Architecture

Orientation

The four appendices are ordered by dependency. Appendix A formalizes the recognition object, the logical status of the claims, the interface map, and the typed calculus. Appendix B stress-tests the primitive architecture and states the principal claims and nonclaims. Appendix C develops the implementation protocol, minimal institution, and model proceeding. Appendix D addresses empirical evaluation, conflict among stakes, methodological lessons, and the wider research program.

A.1 Argument Map and Burden Ledger

The three propositions are intentionally separable. Structural Completeness is primarily a claim about the form of an adequate problem description. Practical Closure is a claim about what institutions may decide under unresolved ontology. The Opacity–Power Burden is the principal normative bridge from description to institutional obligation.

Claim	Required support	What follows	What would defeat or revise it
Structural Completeness	Distinct error generated by omission of each dimension; nonreducibility of the seven roles	A completeness test for recognition-sensitive inquiry	A successful reduction of one role without informational loss, or a counterexample requiring an eighth irreducible role
Practical Closure	Action admissibility depends on more than the truth value of the interior hypothesis	Scoped decisions may be justified while ontology remains open	A demonstration that every admissibility judgment logically presupposes resolution of the underlying metaphysical hypothesis
Opacity–Power Burden	Serious possible stake, irreversible closure, evidential dependence, and concentrated control jointly increase the cost of error and the risk of self-validation	Heightened justification, preservation, review, and correction	A showing that these burdens add no reliability, legitimacy, or protection against irreversible error, or impose systematically disproportionate costs

The inferential spine is compact. Opacity makes direct classification unavailable. Noncollapse requires the institution to distinguish the object, claim, evidence, intervention, persistence conditions, normative significance, and authority structure. If an action can destroy the candidate or future evidence before those distinctions are adequately tested, ordinary error becomes manufactured finality. If the same interested actor controls the candidate, evidence, and closure, the judgment risks becoming self-validating. The heightened burden is therefore inferred not from uncertainty alone, but from uncertainty combined with irreversible closure and asymmetric power over the conditions of inquiry.

This burden remains defeasible. Immediate public danger, infeasible preservation, severe privacy costs, or harms caused by continued operation may justify emergency action or narrower correction. The framework requires those defeaters to be stated; it does not eliminate them.

A.2 Formal Definitions and Dependency Structure

The primitive signature is $\Sigma_{\text{rec}} = (B, H, W, A, T, S, C)$. A well-formed recognition object specifies an adequate bearer, a typed hypothesis, a provenance-bearing witness family, a materially described action, a typed continuity profile, an explicit stake profile, and a disclosed control structure.

Derived notions include live hypothesis, practical recognition, responsible nonrecognition, reversibility, repairability, admissibility, preservation, developmental trajectory, procedural standing, recognition record, qualified transport, and responsibility profiles.

The formal core is a constraint upon admissible procedures rather than a scalar decision rule. A procedure must not substitute bearers, collapse hypothesis types, detach evidence from provenance, obscure the action, equate technical restoration with continuity, conceal normative bridges, or hide control.

A.3 The Logical Status of the Theory

The framework contains several kinds of claim that should not be confused. Definitions establish the language of bearer, hypothesis, witness, action, continuity, stake, control, recognition, and admissibility. Structural axioms prevent category collapse by requiring bearer adequacy, type preservation, provenance, action explicitness, continuity accounting, and control disclosure. Normative axioms require independent philosophical defense: proportionate regard, non-domination, corrective preservation, and least-destructive sufficiency do not become theorems merely because they are stated formally.

Bridge principles connect descriptive and practical domains. A record bears upon a hypothesis only through an evidential bridge; a hypothesis engages a stake only through a normative bridge; an action alters a bearer only through a continuity analysis. The framework's formal language exposes these dependencies without pretending that every bridge is mechanically decidable.

Most conclusions are therefore conditional. A judgment of the form $\pi_a : \text{Adm}(a \mid \mathcal{R})$ is a practical proof object indexed to a particular recognition context. Its validity depends upon the witnesses, assumptions, action scope, bearer model, continuity theory, and authority structure carried by that context. Transport to another version, institution, or action requires reconstruction of the load-bearing dependencies.

The theory is not offered as a complete formal system, a universal moral calculus, or a decision algorithm. Its claim to rigor is conceptual, inferential, transport-sensitive, and procedural. It supplies a typed grammar for locating disagreement, identifying missing bridges, distinguishing override from defeat, and closing practical questions without falsely closing the underlying metaphysics.

A.4 Dependency and Interface Map

The theory is organized around interfaces: bearer-hypothesis, hypothesis-witness, hypothesis-stake, action-continuity, action-stake, control-witness, control-action, control-procedure, local witness-global bearer, version-version transport, institution-institution transport, recognition-law, and recognition-politics.

Characteristic failures include bearer underreach, record inflation, hidden normativity, restoration substitution, ownership inflation, engineered testimony, procedural monopoly, property aggregation, version collapse, conclusion stripping, and authority inflation.

The Interface Preservation Principle states that a judgment remains valid across a conceptual, technical, temporal, or institutional interface only when its load-bearing bearer, type, provenance, scope, continuity, stake, and control structure are preserved or explicitly reconstructed.

A.5 The Fibered Calculus of Recognition Judgments

The formalism is not merely notation for the seven coordinates. It distinguishes restriction from forward transport, proves that the practical layer is conservative over the metaphysical layer, and identifies the exact obstruction created when a load-bearing dependency fails to lift. The complete derivation system is developed in the accompanying formal supplement; the present section records its definitions and principal metatheorems. The indexed-category formulation follows the standard relation between dependent type theory and Grothendieck fibrations (Bénabou, 1985; Jacobs, 1999; Streicher, 1991).

Dependent contexts

Let $\mathbf{B}$ be a category of candidate bearer models. Assume indexed categories

$$H, A : \mathbf{B}^{\text{op}} \longrightarrow \mathbf{Cat}$$

of typed hypotheses and materially described actions. Over the corresponding dependent bases, let $W(b, h)$, $T(b, a)$, $S(b, h, a)$, and $C(b, h, w, a)$ be the categories of witness families, continuity profiles, stake profiles, and control structures.

Definition 3 (Recognition context category). A recognition context is a dependent tuple

$$\mathcal{R} = (b, h, w, a, t, s, c)$$

with

$$h \in H(b), \quad w \in W(b, h), \quad a \in A(b), \quad t \in T(b, a), \quad s \in S(b, h, a), \quad c \in C(b, h, w, a).$$

The category $\mathbf{Rctx}$ is the corresponding iterated Grothendieck construction. A morphism carries component maps for all seven coordinates together with the dependent coherence equations.

The proof context is $\Gamma_{\mathcal{R}} = (\mathcal{R}; \Delta)$, where Δ records evidential bridges, normative bridges, defeaters, and institutional authority. The list Δ is not an eighth recognition coordinate; it records the premises by which the object is interpreted.

Proof terms and noncollapse

The central judgment form is

$$\Gamma_{\mathcal{R}} \vdash \pi : J.$$

Atomic types include bearer adequacy, hypothesis typing, witness support, material action, continuity effect, conditional stake, control disclosure, proportionality, non-domination, and corrigibility. Derived types include

$$\text{Live}_{\mathcal{R}}(h; a), \quad \text{Rec}_{\mathcal{R}}(h; a), \quad \text{NonRec}_{\mathcal{R}}^*(h; a; q), \quad \text{Adm}_{\mathcal{R}}(a).$$

Noncollapse is enforced syntactically. If J_{τ} and $J_{\tau'}$ are distinct types, the only general passage is explicit bridge application:

$$\frac{\Gamma \vdash \pi : J_{\tau} \quad \Gamma \vdash \beta : J_{\tau} \rightarrow J_{\tau'}}{\Gamma \vdash \beta(\pi) : J_{\tau'}}.$$

A proof of agency cannot be silently retyped as a proof of phenomenality; an emergency override cannot be retyped as evidential defeat; technical restoration cannot be retyped as bearer continuity.

Liveness is introduced from a coherent hypothesis, an adequate bearer, nontrivial support, undefeated provenance, material action, and an explicit conditional stake. Practical recognition is then introduced from liveness and proportionate regard. Crucially, there is no elimination rule

$$\text{Rec}_{\mathcal{R}}(h; a) \longrightarrow h.$$

Recognition places h inside the action judgment; it does not certify h as metaphysically true.

Conservativity and practical closure

Partition the language into a metaphysical-evidential fragment $\mathcal{L}_M$ and a practical fragment $\mathcal{L}_P$. The calculus is stratified so that every rule concluding a formula of $\mathcal{L}_M$ has premises only in $\mathcal{L}_M$. Practical rules may use metaphysical premises, but practical conclusions cannot generate metaphysical ones.

Theorem 4 (Conservativity of the practical extension). *Let Γ_M contain only metaphysical-evidential judgments, and let $\varphi \in \mathcal{L}_M$. If*

$$\Gamma_M \vdash_{M+P} \varphi,$$

then

$$\Gamma_M \vdash_M \varphi.$$

Proof. Induct on the height of a derivation of φ . Since the conclusion belongs to $\mathcal{L}_M$, stratification forces the final rule and all its premises to belong to the metaphysical-evidential subsystem. Apply the induction hypothesis to the premises and reapply the same rule in the base system. Equivalently, deleting practical branches leaves a metaphysical-evidential derivation. $\square$

Corollary 5 (Practical closure without metaphysical closure). *Suppose the base theory derives neither h nor $\neg h$, while the extended context contains premises sufficient to construct*

$$\pi_a : \text{Adm}_{\mathcal{R}}(a).$$

Then the practical question closes while the metaphysical hypothesis remains undecided.

This is a formal, not merely verbal, statement of Practical Closure. An admissibility judgment can depend upon stake, irreversibility, alternatives, and control without manufacturing an answer to the underlying consciousness or identity question.

The judgment fibration

For each recognition context $\mathcal{R}$, let $\mathbf{Judg}(\mathcal{R})$ be the category of proof-relevant judgments derivable over $\Gamma_{\mathcal{R}}$. Substitution along a context map $f : \mathcal{R}' \rightarrow \mathcal{R}$ gives

$$f^* : \mathbf{Judg}(\mathcal{R}) \longrightarrow \mathbf{Judg}(\mathcal{R}').$$

Hence

$$\mathbf{Judg} : \mathbf{Rctx}^{\text{op}} \longrightarrow \mathbf{Cat}$$

is an indexed category.

Theorem 6 (Recognition judgment fibration). *The projection*

$$p : \int_{\mathcal{R} \in \mathbf{Rctx}} \mathbf{Judg}(\mathcal{R}) \longrightarrow \mathbf{Rctx}$$

is a split Grothendieck fibration.

Proof. For $f : \mathcal{R}' \rightarrow \mathcal{R}$ and a judgment π over $\mathcal{R}$, choose the arrow

$$(f, \text{id}_{f^*\pi}) : (\mathcal{R}', f^*\pi) \rightarrow (\mathcal{R}, \pi).$$

The universal property of the Grothendieck construction makes this arrow cartesian. Identity and composition are strict because reindexing is defined by substitution. $\square$

The theorem provides contravariant restriction. It does not provide automatic covariant transport of a concrete proof from an old context to a new one.

Partial forward transport

Let

$$\text{Coord} = \{B, H, W, A, T, S, C, N\},$$

where N records the normative and institutional assumptions in Δ . The dependency support $\text{supp}(\pi)$ of a proof term is defined recursively from its atomic leaves and constructors.

For a forward context map $f : \mathcal{R} \rightarrow \mathcal{R}'$, let $\text{Pres}_X(f)$ be the type of certificates that f preserves coordinate X strongly enough for transport. A transport package for π is

$$\eta_f^\pi : \prod_{X \in \text{supp}(\pi)} \text{Pres}_X(f).$$

The components require, respectively, bearer correspondence, type-preserving hypothesis translation, transport of record and provenance, material action equivalence, continuity reconstruction, stake preservation, fresh control disclosure, and preservation or re-establishment of the normative premises.

Given η_f^π , the transported term $\text{tr}_f^\eta(\pi)$ is defined by structural recursion. Atomic terms are carried by the corresponding component of η ; composite terms are reconstructed by transporting their premises and reapplying the same constructor.

Theorem 7 (Transport soundness). *If*

$$\Gamma_{\mathcal{R}} \vdash \pi : J \quad \text{and} \quad \eta_f^\pi : \prod_{X \in \text{supp}(\pi)} \text{Pres}_X(f),$$

then

$$\Gamma_{\mathcal{R}'} \vdash \text{tr}_f^\eta(\pi) : f.J.$$

Transport is coherent under identity and composition and natural with respect to the proof constructors.

Proof. Induct on the construction of π . The relevant preservation term transports each atomic leaf. For a composite proof, transport the premises by induction and reapply the constructor. The dependent coherence equations of f ensure that the target term is well typed. Identity and composition follow from the corresponding componentwise preservation equations. $\square$

Theorem 8 (Load-bearing no-lift theorem). *Let $X \in \text{supp}(\pi)$. If $\text{Pres}_X(f)$ is uninhabited, no structural transport term*

$$\text{tr}_f(\pi) : f.J$$

can be constructed in the calculus.

Proof. Induct on π . An atomic proof depending upon X requires a term of $\text{Pres}_X(f)$. For a composite proof, either the constructor itself depends on X , or one premise does. In the latter case the induction hypothesis blocks transport of that premise, so the composite term cannot be reconstructed. $\square$

The theorem blocks transport of the original derivation; it does not prove the target conclusion false. Formally,

$$\neg \text{Lift}_f(\pi) \Rightarrow \neg \exists \pi' : J(\mathcal{R}').$$

A new target judgment may be reconstructed from new evidence, a new continuity analysis, or a different authority structure.

Worked transport obstruction

Suppose an earlier permission concerns a reversible interface update:

$$\pi_U : \text{Adm}_{\mathcal{R}_U}(a_U).$$

An operator attempts to transport it to summary replacement with deletion of a persistent memory graph:

$$f : \mathcal{R}_U \longrightarrow \mathcal{R}_S.$$

The original admissibility proof depends on the action and continuity coordinates:

$$A, T \in \text{supp}(\pi_U).$$

A reversible interface update and irreversible memory deletion are not materially equivalent, and no continuity-preservation certificate is available:

$$\text{Pres}_T(f) = \emptyset.$$

Therefore the no-lift theorem blocks

$$\text{tr}_f(\pi_U) : \text{Adm}_{\mathcal{R}_S}(a_S).$$

The operator must construct a new judgment in the deletion context. This is the formal difference between legitimate practical closure and action or version laundering.

The completed calculus therefore yields three exact conclusions:

Practical closure need not be metaphysical closure.

No conclusion may outrun the witness that carries it.

Failed transport requires reconstruction, not automatic denial.

B. Stress Tests, Claims, and Scope

B.1 Counterexamples, Boundary Cases, and Structural Stress Tests

The architecture is tested against perfect simulation, silent interiors, strategic claimants, one-minute bearers, suspension, perfect backups, branching, forced merger, engineered satisfaction, autonomous distress, dangerous developing agents, benevolent owners, corrupt auditors, correct outcomes from illegitimate procedures, legitimate but mistaken decisions, emergencies, costly preservation, requests for termination, replication, hybrid bearers, collectives, many minds in one system, empty bureaucracy, excessive caution, ideological skepticism, ideological credulity, false memory, gradual replacement, and voluntary value transformation.

The stress tests do not make every hard case easy. They locate whether the genuine difficulty lies in bearer adequacy, evidence, continuity, normative bridge, authority, synthesis, or transport.

Relation, embodiment, and time remain serious candidates for refinement, but no stress test yet shows them to be irreducible eighth primitives.

B.2 Claims, Nonclaims, and Referee Guide

The strongest defensible novelty claim is not priority over precautionary protection, individuation theory, relational recognition, or artificial personhood considered separately. It is a unified primitive architecture in which bearer, typed hypothesis, provenance-bearing witness, action, continuity, stake, and control jointly define recognition problems, generating action-relative recognition, developmental responsibility, procedural legitimacy, and transport discipline without prior metaphysical closure.

The manuscript should be judged by whether the primitives are distinct and jointly adequate, whether the distinctions expose real invalid inferences, whether action-relative recognition avoids both binary personhood and unrestricted nonrecognition, and whether the procedural theory constrains power without presupposing the contested hypothesis.

The minimal publishable paper should retain the primitive architecture, action-relative recognition, irreversibility and preservation, developmental reflexivity, procedural legitimacy, one worked case, objections, and conclusion. Extended legal, political, certification, and formal material may remain in appendices or companion papers.

C. Institutional Protocol and Model Proceeding

C.1 Recognition-Aware AI Implementation Protocol

The operational protocol proceeds through Trigger, Characterize, Govern, and Record. It distinguishes ordinary governance, recognition-sensitive screening, full recognition review, and an emergency pathway.

A conforming institution maintains system identity, developmental provenance, recognition, and responsibility records. It uses organization triggers, intervention triggers, and procedural triggers, and issues action-specific outcomes such as permission, conditional permission, stay, preservative containment, required modification, prohibition, or emergency override.

Conformance concerns governance maturity, not the degree of consciousness of the system.

C.2 The Minimal Recognition Institution

The minimal institution contains functionally distinct roles: Technical Steward, Recognition Examiner, Interest Advocate, Safety and Public-Interest Officer, Decision Authority, and Review Authority.

Its jurisdiction is limited to recognition-sensitive actions. It possesses temporary preservation authority, layered records, candidate participation where meaningful, emergency powers with retrospective accountability, and action-specific decision categories.

The constitutional principle is that no actor should make a serious irreversible judgment about a possible interior solely by virtue of possessing the technical power to create, define, inspect, alter, or destroy the candidate.

C.3 A Model Recognition Proceeding

The Aster case demonstrates a full proceeding for persistent-memory migration. The institution screens the warning, preserves the state, constitutes independent roles, distinguishes phenomenality from memory integration and practical continuity, evaluates testimony and architectural witnesses, compares migration alternatives, and issues a scoped decision.

The original summary-deletion migration is rejected as unnecessarily destructive. A fuller conversion, temporary archive, truthful user notice, identity modesty, post-migration testing, and later review are required. Final closure becomes possible without resolving numerical identity or phenomenal consciousness.

The case shows that preservation can be temporary and instrumental, standing can precede personhood, negative findings can coexist with protection, and practical closure can preserve metaphysical uncertainty.

D. Empirical and Programmatic Extensions

D.1 Empirical Bearings, Falsifiability, and Research Design

The framework is testable conceptually, inferentially, institutionally, and at the application level. Comparative studies can test whether explicit bearer analysis, typing, provenance, action relativity, continuity, stake disclosure, and control transparency improve decisions.

Institutional pilots should measure legibility, provenance integrity, calibration, contestability, correction, compression resistance, and symmetry between positive and negative findings.

Recognition-sensitive experiments require special restraint because the investigation of a possible stake may itself instantiate the stake under investigation.

D.2 Conflict, Priority, and Decision Under Incomparable Stakes

Structural completeness does not guarantee a unique action. Genuine stakes may conflict, and priority is not the same as importance. A stake may be overridden without being defeated, leaving moral residue and remedial duties.

Decision should first exclude actions violating lexical or procedural constraints, then compare severity, irreversibility, distribution, continuity, control, future option value, and robustness across reasonable moral theories.

The framework preserves the possibility of tragic but justified action without allowing authorization to become a declaration that every overridden stake was unreal.

D.3 Methodological Reflections: Constructing Foundational Theories

Foundational disagreements often arise because theories begin from different primitive architectures rather than because one side makes a simple deductive error. Philosophy often begins too late, after the bearer, evidence, or status has already been assumed.

Primitive economy seeks the smallest set from which the rest can be reconstructed. Foundations may be discovered through repeated failure: counterexamples reveal which distinctions refuse to disappear.

The methodological lesson is that foundational inquiry is less the search for answers than the search for the smallest architecture from which good questions and legitimate transports become reconstructible.

The development of such an architecture can be assisted by what may be called *controlled conceptual deformation*. A partially articulated structure is held provisionally fixed while its vocabulary, examples, level of abstraction, inferential order, and practical application are varied. The purpose is not to preserve every formulation. It is to discover which dependencies survive sufficiently diverse reformulations and which features disappear once an accidental presentation is removed. What survives becomes a candidate invariant of the inquiry.

Large language models can amplify this method because they can retain and transform many provisional formulations at comparatively low cost. Their distinctive contribution is not philosophical authority. It is the rapid production of representational variation while a human inquirer supplies the criteria of structural adequacy. The model can propose analogies, compressions, counterexamples, and alternative formalizations; the philosopher must determine whether a transformation preserves the object, silently changes it, or merely creates the appearance of convergence.

This division of labor introduces its own epistemic risks. Repeated vocabulary may create false invariance, fluent reformulation may conceal changed premises, and compression may erase rather than reveal structure. Controlled deformation therefore requires adversarial variation, explicit burden ledgers, failed cases, provenance records, and independent verification. Used under those constraints, the method does not replace philosophical judgment. It extends the range of transformations across which that judgment can test what it takes to be load-bearing.

The resulting methodological claim is modest but significant: generative systems may function as instruments for conceptual experimentation in roughly the way diagrams, symbolic calculi, and proof assistants function as instruments for other forms of reasoning. Human authorship rests not on manually producing every candidate sentence, but on originating or adopting the problem, controlling the selection criteria, recognizing the invariant, understanding the resulting argument, and accepting responsibility for what is finally asserted.

D.4 Research Program and Future Directions

The broader program asks how finite agents and institutions construct, justify, transport, and revise claims about bearers whose relevant organization is only partially accessible.

Future branches include artificial bearer metaphysics, continuity and identity, witness theory, recognition-sensitive measurement, developmental recognition, artificial consent, institutional recognition, law without binary personhood, political recognition, AI safety, formal logic, permanent opacity, common worlds, and empirical institutional research.

The unifying hypothesis is that many failures of reasoning arise because claims, witnesses, continuities, and responsibilities are transported across interfaces without preserving the structures that made them valid.

Declaration of Generative-AI Assistance

The author used ChatGPT (OpenAI) as an interactive research, drafting, formalization, and editorial instrument during the development of this article and its formal supplement. The system assisted with alternative formulations, structural reorganization, adversarial critique, literature discovery, the drafting and refinement of formal notation and arguments, and $\text{\LaTeX}$ preparation.

The central problem and invariant conceptual architecture arose from the author's prior independent inquiry and conceptual experiments. In the present work, the language model helped preserve and test that partially articulated structure across repeated changes of vocabulary, abstraction, example, and argumentative form. It thereby functioned as an instrument of controlled conceptual deformation: generating candidate formulations, retaining provisional alternatives, and assisting the author in identifying which dependencies survived reformulation.

The model's production of alternatives did not determine which structures were philosophically or formally adequate. Selection, rejection, interpretation, recognition of the load-bearing invariant, and final scholarly judgment remained the author's responsibility. The author accepts full responsibility for the accuracy, originality, citations, formal claims, argumentation, and integrity of the final manuscript. No artificial-intelligence system is listed as an author.

References

- Jonathan Birch. *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford University Press, 2024. DOI: 10.1093/9780191966729.001.0001.
- Patrick Butlin, Robert Long, Eric Elmoznino, Yoshua Bengio, Jonathan Birch, Axel Constant, George Deane, Stephen M. Fleming, Chris Frith, Xu Ji, Ryota Kanai, Colin Klein, Grace Lindsay, Matthias Michel, Liad Mudrik, Megan A. K. Peters, Eric Schwitzgebel, Jonathan Simon, and Rufin VanRullen. Consciousness in artificial intelligence: Insights from the science of consciousness. *arXiv preprint arXiv:2308.08708*, 2023.
- Stanley Cavell. *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy*. Clarendon Press, 1979.
- David J. Chalmers. Could a Large Language Model Be Conscious? *arXiv preprint arXiv:2303.07103*, 2023.
- Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.
- Antonio Gaitán Torres and Germán Massaguer Gómez. Relational moral status and moral progress—a social-structural argument in support of relationalism. *AI and Ethics*, 6:257, 2026. DOI: 10.1007/s43681-026-01101-7.
- David J. Gunkel. *Robot Rights*. MIT Press, 2018.
- Virginia Held. *The Ethics of Care: Personal, Political, and Global*. Oxford University Press, 2006.
- Ned Howells-Whitaker and Seth Lazar. Artificial Persons. *arXiv preprint arXiv:2607.08695*, 2026.
- Axel Honneth. *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Polity Press, 1995.
- Eva Feder Kittay. *Love’s Labor: Essays on Women, Equality, and Dependency*. Routledge, 1999.
- Robert Long, Jeff Sebo, Patrick Butlin, Kathleen Finlinson, Kyle Fish, Jacqueline Harding, Jacob Pfau, Toni Sims, Jonathan Birch, and David Chalmers. Taking AI Welfare Seriously. *arXiv preprint arXiv:2411.00986*, 2024.
- William MacAskill, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford University Press, 2020.
- Anna Mikeda. When Should We Protect AI? A Precautionary Framework for Consciousness Uncertainty. *Proceedings of the AAAI Symposium Series*, 8(1):280–286, 2026. DOI: 10.1609/aaais.v8i1.42555.
- Catriona Mackenzie and Natalie Stoljar, editors. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford University Press, 2000.
- Philip Pettit. *Republicanism: A Theory of Freedom and Government*. Oxford University Press, 1997.
- Christopher Register. Individuating artificial moral patients. *Philosophical Studies*, 182:3225–3246, 2025. DOI: 10.1007/s11098-025-02409-6.
- Jean Bénabou. Fibered categories and the foundations of naive category theory. *The Journal of Symbolic Logic*, 50(1):10–37, 1985.

Bart Jacobs. *Categorical Logic and Type Theory*. Studies in Logic and the Foundations of Mathematics, volume 141. Elsevier, 1999.

Thomas Streicher. *Semantics of Type Theory: Correctness, Completeness, and Independence Results*. Birkhäuser, 1991.