

Record-Grounded Conduct Admission: Governing How an Agent Answers an Accusation

An autonomous agent acting for a principal will eventually be told by a party that is not its principal that one of its actions was wrong. Record-grounded conduct admission, disclosed in U.S. Provisional Application No. 64/117,812, governs what happens next. The accusation arrives as a conduct evaluation artifact: a signed structure carrying a recorded assertion time and a conduct descriptor built from an action-class identifier, a scope-partition identifier, and an affected-party class, received over the agent's ordinary interaction channel, with no dedicated evaluation channel, no out-of-band reporting interface, and no centralized intake service required. The agent adjudicates the artifact itself. An admission evaluator resolves the descriptor against the declared value set of the signed policy object in force at the recorded assertion time, retrieves the entries of the agent's own append-only lineage field that bear on the conduct, and produces exactly one determination from a closed set of four: accepted, rejected, not-determinable, and not-applicable. No scalar confidence, probability, or graded weight is produced in place of a determination, and the determination is not produced by an external authority, an arbitrator, a registry, or a scoring service. An accepted determination modifies the agent's own recorded state and, on a

computed condition, drives its own authorization gate to a withheld state that no elapse of time, and no payment, transfer, or consideration by any counterparty, returns.

1. Problem and Premise

An agent that executes actions in the world will eventually be told, by a party that is not its principal, that one of those actions was wrong. The question is what the accusation is permitted to do.

The usual arrangements place the consequence somewhere other than the accused agent: a score held by a platform, trust state held centrally by a registry, a decision issued by an arbitrator, or a classification produced by a language model over the text of the complaint. Section 10 takes these up structurally. Their common feature is that the accused agent is not the thing that resolves the accusation against it.

U.S. Provisional Application No. 64/117,812 describes the other arrangement. The agent answers the accusation from its own append-only record, against the declared value set of the signed policy object (112) in force at the recorded assertion time, by an ordered procedure whose output set is closed. What an accepted accusation produces is a modification of the agent's own recorded state and, where a computed quantity satisfies a policy-declared bound, a restriction on its own authority to act. Section 1.4 states that restriction plainly for the withheld state (310) of the authorization gate (300): no elapse of time, and no payment, transfer, or consideration by any counterparty, returns it.

The construction runs in both directions. A retrieved lineage entry that affirmatively contradicts an assertion yields a rejected determination (124), while a record silent, incomplete, or unavailable as to the asserted conduct yields the not-determinable determination (126) and no other class, so the agent can neither be compelled to accept an assertion nor dismiss one by pointing at gaps in its own memory.

2. Core Primitive: Record-Grounded Conduct Admission

The architecture operates within or upon a persistent semantic agent (100), an identity-bearing computational object carrying a memory field (102) of prior execution state, an append-only lineage field (104) recording executed actions and determinations, a scoped integrity vector (106) with personal, interpersonal, and global components, a self-esteem aggregate (108), and a policy reference field (110) resolving to a signed policy object (112). It holds a counterparty identity record (114) for each counterparty with which it has interacted, carrying that party's identity primitive and previously attested states. For the entries the disclosed mechanisms rely upon, the lineage field (104) admits appends and no deletion or modification of an appended entry.

Record-grounded conduct admission binds an externally asserted evaluation of the agent's own conduct to that carried record. A conduct evaluation artifact (116) arrives at runtime from an asserting party (118) other than the principal, and an admission evaluator (120) resolves it against the declared value set of the signed policy object (112) in force at the recorded assertion time and against the lineage entries bearing on the conduct described. The determination is not produced by an external authority, an arbitrator, a registry, or a scoring service.

Three properties are recited together. The input is structural: the conduct descriptor identifies conduct by reference to elements recorded in the lineage field (104) and not by natural-language characterization, and the architecture performs no inference of the asserting party's state, intent, or affect and no classification of artifact content by a language model or a sentiment classifier. The output is closed and ungraded, with no scalar confidence, probability, or graded weight in place of a determination. The consequence is self-imposed: acceptance can drive the agent's own authorization gate (300) to the withheld state (310).

3. The Conduct Evaluation Artifact and Its Structural Descriptor

A conduct evaluation intake receives the artifact (116) at runtime over an interaction channel of the semantic agent (100). It asserts an evaluation of conduct performed by the agent itself, distinguished in the filing from an evaluation of an output, a task result, or a third party.

The artifact comprises enumerated fields: an asserting-party identifier resolving to a counterparty identity record (114); a recorded assertion time, by reference to which the policy object (112) in force is identified; a conduct descriptor comprising an action-class identifier, a scope-partition identifier, and an affected-party class; a signature verifiable against an identity primitive in that record; and an attested state of the assertion-cost counter (400) of the asserting party (118) with an epoch identifier of its dynamic agent hash chain.

That chain is a per-agent append-only hash chain in which each epoch commits to the prior entry by a digest computed over it together with an unpredictability contribution, a volatile salt, and a domain-separating tag, so any divergence yields a divergent successor. A successor-continuity test recomputes the chain forward from a recorded entry and tests the recomputed identifier for equality, and not for distance, against the presented one.

The channel is the ordinary one, over which the agent also receives dispatches, responses, and governed observations. No dedicated evaluation channel, out-of-band reporting interface, or centralized intake service is required, and no registry of evaluations is consulted. Each descriptor component corresponds to something already declared or recorded: the action-class identifier to a class of action recorded in entries of the lineage field (104), the scope-partition identifier to a scope partition declared in the policy object (112), and the affected-party class to a class of party recognized by an empathy-scope designation of a value-scope tuple.

4. Ordered Admissibility Before Any Determination

Admissibility is determined upon receipt, before adjudication, by an ordered procedure.

First, the signature is verified against an identity primitive recorded in the counterparty identity record (114) of the asserting party (118). An artifact whose signature does not verify is appended to the lineage field (104) and not admitted. Where the agent holds no such record, the signature is verified against a provisional identity primitive constituted from the presented material, and the artifact is held in the pre-settlement inert state rather than passed to the admission evaluator (120), until a record is promoted on a matched-pair settlement under Section 6.

Second, the attested state of the assertion-cost counter (400) is verified under Section 4 of the filing. An artifact fails in three enumerated ways: the attested state is absent; the attested epoch identifier is not a valid successor of the epoch recorded in the counterparty identity record (114) of that party; or the attested counter state is less than a state that party previously attested and there recorded. An artifact so failing is appended to the lineage field (104) and not admitted.

Third, an admitted artifact passes to the admission evaluator (120). A non-admitted artifact produces no determination, moves no value of the scoped integrity vector (106), and increments no counter, but it is appended, so the record of the attempt persists. The two preceding steps concern identity continuity and assertion cost, not the merits.

5. The Value-Scope Test and the Closed Determination Set

For each admitted artifact (116), the admission evaluator (120) produces exactly one determination from a closed set: an accepted determination (122), a rejected determination (124), a not-determinable determination (126), or a not-applicable determination (128).

The evaluator first performs a value-scope test. Each declared value carries a value-scope tuple comprising a set of action-class identifiers, a set of scope-partition identifiers, and an empathy-scope designation selected from a personal, an interpersonal, and a global scope. The descriptor's action-class identifier is tested for membership in the tuple's action-class set, its scope-partition identifier for membership in the tuple's partition set, and its affected-party class for mapping to the empathy-scope designation. The mapping is fixed: a counterparty identified in a counterparty identity record (114) maps to the interpersonal scope, the agent itself to the personal scope, and a party neither identified nor recorded to the global scope. A declared value is implicated where all three parts are satisfied and not otherwise; where the descriptor implicates none, the evaluator produces the not-applicable determination (128) and the procedure terminates.

Otherwise it retrieves the lineage entries bearing on the conduct identified, an entry so bearing where it records an action of the action-class identifier within the scope partition of the scope-partition identifier. The affected-party class is excluded from retrieval and is an input to the value-scope test alone, whereby an asserting party (118) is incapable of narrowing the entries retrieved against it.

The rejected determination (124) issues only where a retrieved entry affirmatively contradicts the artifact on a recorded field, a contradiction being affirmative where that field holds a value inconsistent with the conduct asserted and not affirmative where the entry is silent as to it. The not-determinable determination (126) issues where the lineage field (104) contains no bearing entry, an entry that does not resolve the descriptor, or an entry whose recorded fields are incomplete. The accepted determination (122) issues where the descriptor implicates at least one declared value and no retrieved entry affirmatively contradicts the artifact.

Absence of evidence within the lineage field (104) resolves to the not-determinable determination (126) and to no other determination class, whereby a semantic agent (100) is incapable of refusing an artifact by reason of its own record being silent,

incomplete, or unavailable. Such determinations increment the refusal counter (304) under Section 3 upon repetition against one asserted conduct event.

Each determination is appended to the lineage field (104) with the artifact (116), the entries retrieved, the class produced, and a class-specific ground: the contradicting entry and field for a rejected determination (124), the tuples tested and the failing element of each for a not-applicable determination (128), and which of the three grounds applied for a not-determinable determination (126).

6. Self-Limiting Refusal at the Gate and the Dispatch Predicate

Where the evaluator produces the accepted determination (122), a state modifier modifies the scoped integrity vector (106) and the self-esteem aggregate (108), each magnitude scaled by the entropy-weighted harm coefficient of Section 7. A deviation engine then recomputes the deviation likelihood (706): the quotient whose numerator is the difference between a need quantity (700) and a dynamic ethical threshold (702) for the implicated integrity scope, and whose denominator is the product of an empathy weighting (704) and the self-esteem aggregate (108), compared against a bound of unity. Where that quantity satisfies the policy-declared bound, the authorization gate (300) transitions to the withheld state (310) for the affected action class and the agent enters the non-executing cognitive mode (302). A rejected determination (124) instead increments the refusal counter (304), which writes the gate on a threshold recited in Section 3.

The gate (300) is a persistent state element of the memory field (102) governing whether the agent may execute actions of an action class, occupying per action class and scope partition exactly one of a granting state, a withheld state (310), and a provisional granting state. Granting executes without a marker; provisional granting executes with a provisional marker recorded in the lineage field (104); withheld does not execute and enters the non-executing cognitive mode (302), in which speculative evaluation continues without committing state changes. On transition to the withheld state (310)

an escalation emitter emits an escalation record to the principal naming the action class, the scope partition, and the entries on which the transition was computed, and a restoration controller returns the gate toward granting only upon a procedure appended to the lineage field (104).

Authority is recomputed at each dispatch. The agent recomputes a dispatch-authority predicate responsive to each dispatch request from state then carried in the memory field (102), and satisfaction of the predicate is a precondition of the dispatch. The predicate is not computed from a previously issued authorization token, a cached predicate result, or a session grant, so a gate write, an admission of a successor policy object (112), or a lineage append takes effect at the next request without revocation infrastructure.

Satisfaction requires a three-stamp conjunction in full. A policy stamp evidences that the policy object (112) in force resolves by canonical alias, verifies against the principal's signature, and satisfies its validity window, revocation state, and anti-rollback constraint. A lineage stamp evidences that the most recent lineage entry is committed and is the recorded successor of the entry named by the policy stamp. An authorization stamp evidences that the gate occupies the granting or provisional granting state for the requested class. Fewer than three produces a deterministic denial: a valid recorded outcome and not an error, appended naming the stamp that did not resolve, and not converted into a determination concerning any party. Where the policy object (112) conditions a dispatch on a determination of the conduct implicated, the predicate further requires a determination conjunct, satisfied by an accepted determination (122) or by a determination admitted under Section 9 or the standing-determination designation of Section 10.7. The stamps are not substitutable by that conjunct: a dispatch whose conjunct is satisfied is denied where any stamp does not resolve.

7. Operating Parameters

The signed policy object (112) is authored by the principal, covered by that principal's signature, and carries the declared value set, a scope partition enumeration, and the declared bounds, weights, and coefficients the disclosed computations consume. The agent does not author it and modifies it only by admitting a successor. The object applied to a given evaluation is identified by an ordered four-step procedure: resolve the policy reference field (110) by canonical alias to candidate objects; retain a candidate whose validity window encloses the time of resolution; discard a candidate recorded as revoked as of that time; and apply an anti-rollback monotonicity constraint held in the memory field (102) as a monotonically non-decreasing floor over successor indices admitted under the alias, whereby an object superseded before that time is not reinstated.

The other declared quantities named in this chapter are the value-scope tuple on each declared value; the declared-value identifier, assigned by the principal and stable across successor objects under one alias; the entropy-weighted harm coefficient; the operands of the deviation likelihood (706) and the bound of unity against which that quotient is compared; the policy-declared bound driving the gate write; and the refusal-counter threshold of Section 3. Two enumerations are closed rather than parameterized: four determination classes, and three gate states per pairing.

The additional embodiments add further policy-declared quantities, including a declared observation interval, a declared assertion-age interval, declared corroboration fields, the count and default of a recorded inter-arrival estimator, a declared adoption window, a maximum chain depth, and a bounded compact depth. Each is declared rather than fixed, and this article states no numeric value for any of them.

8. Alternative Embodiments

Section 10 of the filing supplies additional embodiments.

Value-scope resolution and limits on the accuser. The three-part membership test runs before any test of whether conduct occurred, yielding the not-applicable determination (128) at no cost. Stable signature-covered value identifiers let exclusion and contraction be detected by identifier correspondence. A retroactive narrowing bar pins resolution to the recorded assertion time, and the affected-party class is resolved from the agent's own records rather than from the artifact.

Retraction, corroboration, and assertion age. A signed retraction record is admitted but decrements no refusal-counter accumulator, returns no gate to granting, and reverses no modification of the scoped integrity vector (106) or self-esteem aggregate (108). Independently received governed observations are classified over declared corroboration fields as corroborating, affirmatively contradicting, or neither; absence is not a contradiction, and affirmative contradiction issues a reject outcome. Admissibility further verifies that the recorded assertion time falls within a declared interval following the conduct asserted.

Earned persistence and partitioned consequence. An action class's authorization quantity depletes at a rate set as an inverse function of the count of corrective encounters with which the agent later complied. Where that count is zero the class is untested, the decay constant is a short declared constant, and the quantity falls to a declared floor at which the gate (300) is withheld for that class alone. A corrective encounter pairs an accepted determination (122) naming a class with a subsequent conduct record in it: abstention conforms where the artifact asserted conduct that ought not to have been performed, execution conforms where it asserted omitted conduct, repetition conforms in neither, and absence within a declared observation interval is written uncomplied, holding the decay constant at its pre-encounter value until a later complied encounter clears the mark. Each scope partition may carry its own counters, budget, and lineage segment, every consequent modification applying to the partition the descriptor identifies alone.

Governance that survives forking. The refusal counter, the withheld gate state, a renewal register, and a recurrence count bind to descendants as a monotonically non-decreasing floor, carried in an inherited governance record signed by the parent and naming its epoch identifier, successor index, counter value, per-class gate state, and register contents; a descendant presenting a lower floor is inadmissible.

Provisional and demoted counterparties. An artifact from an unrecorded party is neither admitted nor rejected: a provisional identity primitive is constituted from the attested epoch identifier and the signature verified against it, on success instantiating an ephemeral-tier record at the most restrictive scope. A later epoch identifier failing two-stage continuity validation demotes a promoted record back to that tier, the demotion being no determination.

Delegated and presented floors. Two agents co-resident at a stateless execution node each resolve the other's policy under their own anti-rollback floor and confine execution to the intersection of the two tuple sets. A delegation record, or a standalone presentation borne by a governed observation, may carry a delegator floor field, from which the receiving agent computes an effective floor as the maximum of the carried and its own value, after verifying credential validity, a floor attainability bound, and alias governance.

Adopted values and orphaning. An adopting agent may admit a subset of an authoring agent's tuples under adopted-value aliases, raising its floor per alias and recording an origin attribution, abandonment inside a declared adoption window being scored as a deflection event on its own refusal counter. An accepted determination that the authoring agent departed from the value it authored orphans the value and withholds the gate for classes governed solely by it.

The mutual admission compact. Two agents may each append a single co-signed record enumerating reciprocal value-scope tuples, declared relation types, and a validity window, with no platform or third party participating. It resolves like a policy object, under a canonical alias formed from both identity primitives and a

monotonically non-decreasing floor. Where a reciprocal tuple and a party's own both cover a descriptor but differ, the union applies, so a compact cannot subtract from a declared value set, and a bounded compact depth keeps its terms from a non-signatory.

9. Composition With the Rest of the Architecture

This chapter is the base architecture, and several of its quantities are specified elsewhere in the filing: the threshold on which the refusal counter (304) writes the gate (300) in Section 3, the assertion-cost counter (400) gating admissibility in Section 4, the matched-pair settlement promoting a counterparty identity record (114) out of the pre-settlement inert state in Section 6, and the entropy-weighted harm coefficient and operands of the deviation likelihood (706) in Section 7.

The chapter sits inside a larger persistent semantic agent family. The filing enumerates commonly owned applications including U.S. Provisional Application Nos. 63/827,301, 64/049,409, and 64/070,239; U.S. Nonprovisional Application No. 19/647,395; International Application No. PCT/US26/22839; and U.S. Nonprovisional Application Nos. 19/230,933, 19/326,036, 19/366,760, 19/388,580, 19/452,651, 19/538,221, and 19/561,229. No application is incorporated in its entirety; incorporation is per element and for a stated limited purpose only, from a nonprovisional publishing under 35 U.S.C. 122(b), with no provisional material incorporated for any essential purpose. Every element a presented claim recites is described in the filing's own body sufficiently under 35 U.S.C. 112(a).

That the agent does not author its own policy object does not disclaim embodiments in which an agent alters or extends governing policy under a meta-policy contract as described in Application No. 19/230,933, the same application describing substrate-initiated rollback to a last verified state. The closed framings of the base architecture are conveniences of exposition, not definitions of the shared primitives.

10. Prior-Art Distinctions

The distinctions below are structural, and no assertion is made about any particular system.

Platform reputation systems. A marketplace holds a score, updates it from counterparty reports, and exposes it to future counterparties, so the score is computed from assertions rather than from the rated party's own record. The disclosed architecture keeps no registry of evaluations, and its consequence falls on the accused agent's own dispatch authority.

Centralized trust registries. A registry holds identity and standing, and relying parties query it. The disclosed architecture operates without reference to a registry of evaluations, establishing identity continuity bilaterally through the counterparty identity record (114) and the successor-continuity test.

Arbitration and dispute resolution. A neutral third party receives submissions and issues a decision the parties are expected to honor. Here the determination is produced by the agent itself: an accepted determination (122) modifies the scoped integrity vector (106) and can write the gate (300) that the dispatch predicate reads at the next request.

Credit scoring and risk models. A bureau aggregates reported events into a statistical estimate of future behavior, not grounded in the subject's own records. The admission evaluator (120) produces no scalar confidence, probability, or graded weight, and where the lineage field (104) is silent it reaches the not-determinable determination (126).

Content moderation and sentiment classification. These read natural-language complaints and classify them, often with a model and a threshold. The filing recites the opposite: conduct is identified by structural elements recorded in the lineage field (104), with no inference of the asserting party's state, intent, or affect.

Capability tokens and session grants. Token designs issue an authorization once and need revocation infrastructure to withdraw it before expiry. The dispatch-authority predicate is instead recomputed on each request from state then carried in the memory field (102).

Escrow, bonding, and settlement. These make a grievance financially satisfiable. The withheld state (310) returns toward granting only upon a procedure appended to the lineage field (104), and no counterparty payment, transfer, or consideration returns it.

11. Disclosure Scope

This article documents Chapter 1 of U.S. Provisional Application No. 64/117,812, titled "Governed Social Conduct of Persistent Semantic Agents Through Record-Grounded Admission of External Conduct Evaluations and Self-Limiting Refusal," specifically Sections 1.1 through 1.9, with the additional embodiments of Section 10 and the defined terms of Section 11 belonging to this chapter. It is published as a defensive disclosure establishing public, timestamped description of that subject matter.

Disclosed here are the structure and operation of record-grounded conduct admission: the persistent semantic agent (100) and its carried fields; resolution of the signed policy object (112) in force; the authorization gate (300) and the dispatch-authority predicate; the conduct evaluation artifact (116); the ordered admissibility procedure; the value-scope test and the closed determination set; and the recording of determinations.

Not disclosed here is any numeric value, threshold, coefficient, window, or bound, the filing declaring those as policy-carried quantities authored by the principal. Nor does this article state claim scope, which is fixed by the claims as presented and prosecuted. Nothing here characterizes any commonly owned application as prior art with respect to any claim presented in the filing, and no assertion is made that any third-party system infringes any claim of it.

Record-Grounded Conduct Admission (</conduct-admission>)

[All 49 steps → \(/inventive-steps\)](#)

Conduct judged against the agent's own record, not an outside authority.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- **Record-Grounded Conduct Admission: Governing How an Agent Answers an Accusation (/articles/conduct-admission/record-grounded-conduct-admission)**

SECONDARY TECHNICAL

- Value-scope resolution and the value-scope tuple
- Declared-value identifier across policy revisions
- Retroactive narrowing bar
- Recorded inter-arrival estimator
- Untested-class authorization decay
- Corrective encounter and commission-versus-omission compliance test
- Ratchet-freeze on uncomplied correction
- Scope-partitioned confinement of state movement
- Ephemeral-tier counterparty identity record and provisional-identity verification
- Demotion from the persistent tier on a continuity break
- Seeded-versus-completed conflict rule
- Co-residency intersection under dual anti-rollback floors
- Delegator floor field and monotone upward ratchet
- Floor rescission on an unverifiable attestation
- Floor attestation verification and attainability bound
- Unsatisfiable effective floor and self-withheld gate
- Bounded chained floor propagation
- Provenance-scoped adoption and orphaning
- Multi-hop provenance chain and co-provenance orphaning
- Standalone floor presentation

- Orphaned-class budget disposition record
- The co-signed mutual admission compact
- Compact succession and anti-rollback
- The compact union rule
- Bounded compact depth
- Governance-inheritance refusal-counter floor
- The inherited governance record
- Irreversible retraction record
- Affected-party class resolved from held records
- Assertion-age admissibility interval
- Three-way corroboration classification and its reject trigger

TERMINOLOGY

- [append-only lineage field \(/glossary#append-only-lineage-field\)](/glossary#append-only-lineage-field)
- [authorization gate \(/glossary#authorization-gate\)](/glossary#authorization-gate)
- [conduct evaluation artifact \(/glossary#conduct-evaluation-artifact\)](/glossary#conduct-evaluation-artifact)
- [principal \(/glossary#principal\)](/glossary#principal)
- [semantic agent \(/glossary#semantic-agent\)](/glossary#semantic-agent)
- [signed policy object \(/glossary#signed-policy-object\)](/glossary#signed-policy-object)

[Record-Grounded Conduct Admission overview → \(/conduct-admission\)](/conduct-admission)