Nick Clark · Published July 29, 2026 · PDF

Intercom Fin: Resolution Rates and the Conduct Record

A customer writes back after an automated support conversation and says the agent behaved badly, not that it answered wrongly. Customer service AI agents such as Intercom Fin are, as publicly described, instrumented around the quality and disposition of answers: responses grounded in the operator's own content, reporting on conversations resolved, and handoff to a human teammate. That instrumentation is well matched to the job the category defines. Record-grounded conduct admission, disclosed in Chapter 1 of U.S. Provisional Application No. 64/117,812, takes a different object: a signed assertion about the agent's own conduct, arriving from a party other than the principal, resolved by the agent against its own append-only lineage field and the signed policy object in force into exactly one of four determination classes, and not by an external authority, an arbitrator, a registry, or a scoring service.

When the Complaint Is About the Agent, Not the Answer

A customer closes a support conversation and writes back the next day. The AI agent, they say, told them a refund was approved, then treated the same request as ineligible, citing a condition nobody had disclosed. That is not a report that an answer was wrong.

It is an assertion about how the agent behaved, made by someone who is not the operator, and it lands in the same inbox as every routine question.

Support organizations have mature machinery for the first kind of complaint: answers get reviewed, conversations get audited, unsatisfied customers get routed to a person. The second kind sits differently. The party holding the record of what the agent did is the agent itself, while the asserting party has no access to the operator's logs. Familiar approaches place a further party between the assertion and the record, whether a reviewer, a scoring service, or a rating attached to the vendor. The structural question is narrow: when someone other than the principal asserts that an agent behaved in a particular way, what resolves the assertion, against what record, and into what outcome?

Fin and the Support-Agent Category

Intercom is a customer communication platform, and Fin is its AI agent for customer support. As publicly described, Fin answers customer questions by drawing on the operator's own support content and connected sources, hands off to a human teammate when the conversation should not continue automatically, and reports to the operator on conversations it resolved. Operators configure the scope within which it answers and acts.

That outline sketches a category as much as one product, and the category is well engineered for the objective it addresses. Each of those publicly described elements is instrumentation over an output or a disposition: whether an answer traces to a source, whether the problem went away, when a person takes over. The architecture below takes a different object as input, one the filed disclosure defines by exclusion: an assertion of conduct performed by the agent itself.

Resolving an Assertion Against the Agent's Own Record

Chapter 1 of U.S. Provisional Application No. 64/117,812 discloses a governed social conduct architecture operating within a persistent semantic agent (100). The agent carries a memory field (102), an append-only lineage field (104) recording executed actions and determinations, a scoped integrity vector (106), a self-esteem aggregate (108), and a policy reference field (110) resolving to a signed policy object (112) authored and signed by the principal to which the agent is bound. The agent does not author that object; it modifies it only by admitting a successor.

The input is a **conduct evaluation artifact (116)**, received over the agent's ordinary interaction channel and originating from an asserting party (118) other than the principal. No dedicated evaluation channel, no out-of-band reporting interface, and no centralized intake service is required.

The artifact carries an enumerated field set: an asserting-party identifier resolving to a counterparty identity record (114); a recorded assertion time, identifying the signed policy object in force; 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 party's assertion-cost counter (400) with an epoch identifier of its hash chain.

That descriptor identifies conduct by reference to structural elements recorded in the append-only lineage field (104), and not by natural-language characterization. The architecture performs no inference of the asserting party's state, intent, or affect from the artifact's content, and no classification of that content by a language model or a sentiment classifier. It is consumed by reference to its enumerated fields alone.

Admissibility runs as an ordered procedure. The signature is verified first against an identity primitive in the counterparty identity record (114). The attested assertion-cost counter state is verified second, failing where it is absent, where the attested epoch identifier is not a valid successor of the epoch recorded for that party, or where it falls

below a state that party previously attested. An artifact failing either step is appended to the lineage field and not admitted, producing no determination, moving no value of the scoped integrity vector (106), and incrementing no counter.

The admission evaluator (120) produces, for each admitted artifact, 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)**. Nothing outside that set is produced, and no scalar confidence, probability, or graded weight stands in place of a determination. The determination is produced by the agent over its own lineage field and against the declared value set of the signed policy object (112) in force at the recorded assertion time, and not by an external authority, an arbitrator, a registry, or a scoring service.

A value-scope test runs first. The action-class identifier, the scope-partition identifier, and the affected-party class are each tested against the value-scope tuple of a declared value, which is implicated only where all three parts are satisfied. Where the descriptor implicates no declared value, the not-applicable determination (128) is produced and the procedure terminates. Otherwise the evaluator retrieves lineage entries bearing on the identified conduct. The affected-party class is no input to that retrieval, so an asserting party cannot narrow the entries retrieved against its assertion by the class it declares.

Three outcomes follow from what those entries hold. The rejected determination (124) is produced only where a retrieved entry affirmatively contradicts the artifact on a recorded field, an entry silent as to that field not being an affirmative contradiction. The not-determinable determination (126) is produced where the lineage field holds no bearing entry, an entry that does not resolve the descriptor, or an entry incomplete with respect to it. The accepted determination (122) is produced where a declared value is implicated and no retrieved entry affirmatively contradicts the artifact. Absence of evidence resolves to not-determinable and to no other class, so an agent cannot refuse an artifact by reason of its own record being silent, incomplete, or unavailable.

Each determination is appended to the lineage field with the artifact, the entries retrieved, the class produced, and its ground, which for a rejected determination names the contradicting entry and the field carrying the contradiction.

Two Questions That Do Not Substitute for Each Other

The convergence is genuine: both concern an autonomous agent facing customers without a person in the loop, and what should happen when that goes badly. The divergence is structural, and it shows when what the disclosure requires is set beside what the category's public description covers.

- **Object of evaluation.** The disclosure requires an input asserting conduct performed by the agent itself, expressly distinguished from an evaluation of an output, a task result, or a third party. Publicly described support-agent instrumentation is directed at answers and conversation outcomes.
- **Where resolution happens.** The disclosure requires the determination to be produced by the agent over its own lineage field (104) and against the signed policy object (112) in force at the recorded assertion time, and not by an external authority, an arbitrator, a registry, or a scoring service. Resolution reporting, publicly described, is a measure surfaced to the operator.
- **Shape of the output.** The disclosure requires exactly one determination per admitted artifact from a closed set of four, and forbids a confidence, a probability, or a graded weight in its place. A resolution measure is continuous and aggregated across conversations.
- **What follows.** Handoff to a human teammate is the publicly described path when a conversation should not continue automatically. Under the disclosure, where the accepted determination (122) is produced, a state modifier modifies the scoped integrity vector (106) and the self-esteem aggregate (108), each modification scaled by the entropy-weighted harm coefficient, and a deviation engine recomputes the deviation likelihood (706). 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). Where the rejected determination (124) is produced, a refusal counter (304) is incremented, and that counter writes the gate upon satisfaction of a threshold.

The positioning is complementary. Nothing in the disclosed architecture answers a customer question, and a conduct determination displaces no part of a support agent's answer-quality instrumentation.

Coexisting, and What the Filing Does Not Address

Layer the two in one deployment. A support agent of the publicly described kind handles the customer interaction, grounded in the operator's content and escalating to a person under the operator's configuration. Underneath, each governed action it dispatches is recorded in the lineage field (104) as an action of an action class within a scope partition, and a counterparty's assertion about its conduct arrives as a conduct evaluation artifact (116) over the channel the agent already uses.

Resolution from there is local. The agent verifies admissibility, produces one of the four determinations, and appends it with its ground. Upon a transition of the authorization gate (300) to the withheld state (310), an escalation record naming the action class, the scope partition, and the entries on which the transition was computed goes to the principal, and the agent stops executing that class while speculative evaluation continues without committing state changes. A restoration controller returns the gate toward the granting state only upon a procedure appended to the lineage field; no elapse of time, and no payment, transfer, or consideration by any counterparty, returns it.

The limits of the filed material deserve equal plainness. Nothing in it assesses whether an answer was correct or well sourced, computes a resolution measure, or decides when a conversation should go to a human. Its intake is defined over signed artifacts carrying structural descriptors and counter attestations, a different shape from an unsigned

free-text complaint typed into a chat window, and converting one into the other is deployment work the filing does not perform. The architecture operates without reference to a registry of evaluations. The deviation bound and the refusal threshold are policy-declared by the principal, stated as declared rather than fixed. And an agent whose record is silent or incomplete as to the asserted conduct produces not-determinable determinations (126), which upon repetition against one asserted conduct event increment the refusal counter (304).

Disclosure Scope

This article describes subject matter disclosed in Chapter 1 of U.S. Provisional Application No. 64/117,812, a pending application, and is published as a dated public disclosure of that architecture. No claim scope is asserted here.

References to Intercom Fin are to public materials and are used for comparison only; no relationship, endorsement, or infringement is asserted. The comparison above is architectural and asserts nothing about any product's internals.

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: Testing Whether a Declared Value Is Even Implicated (/articles/conduct-admission/value-scope-resolution)
- The Retroactive Narrowing Bar: Why an Agent Cannot Edit Its Values After the Fact (/articles/conduct-admission/retroactive-narrowing-bar)
- The Delegator Floor: A Monotone Upward Ratchet on Delegated Authority (/articles/conduct-admission/delegator-floor-ratchet)
- The Co-Signed Mutual Admission Compact: Two Agents Binding Each Other (/articles/conduct-admission/co-signed-mutual-admission-compact)
- The Inherited Governance Record: Fork- and Clone-Resistant Accountability (/articles/conduct-admission/inherited-governance-record)
- The Corrective Encounter: A Commission-Versus-Omission Compliance Test for Whether an Agent Actually Changed (/articles/conduct-admission/corrective-encounter-commission-omission-compliance-test)
- Continuity-Break Demotion of a Counterparty Identity Record: Reversing First-Encounter Promotion Without a Conduct Determination (/articles/conduct-admission/demotion-persistent-tier-continuity-break)
- Ratchet-Freeze on Uncomplied Correction: Why One Conforming Act Cannot Restore an Agent's Earned Authorization Persistence (/articles/conduct-admission/ratchet-freeze-uncomplied-correction)
- Three-Way Corroboration Classification: Corroborating, Contradicting, or Neither, and the Failed-Corroboration Reject Trigger (/articles/conduct-admission/three-way-corroboration-classification-reject-trigger)
- Untested-Class Authorization Decay: Metering Agent Permission Persistence by Recorded Corrective Encounters (/articles/conduct-admission/untested-class-authorization-decay)

APPLICATIONS · GENERAL

- Agent Marketplaces: Handling a Complaint Against an Autonomous Seller (/articles/conduct-admission/agent-marketplace-dispute-handling)
- Clinical AI Agents: Answering an Accusation Without a Central Registry (/articles/conduct-admission/clinical-agent-accountability)

APPLICATIONS · SPECIFIC

- The EU AI Act and Inter-Agent Conduct: Where Record-Keeping and Oversight Duties Stop (/articles/conduct-admission/eu-ai-act-accountability-records)
- Sierra AI Agents and the Record Behind a Customer Complaint (/articles/conduct-admission/sierra-ai)

- Decagon: Support Agents and Record-Grounded Accountability (/articles/conduct-admission/decagon-ai)
- **Intercom Fin: Resolution Rates and the Conduct Record (/articles/conduct-admission/intercom-fin)**
- Zendesk AI Agents: Escalation, Audit, and Agent Conduct (/articles/conduct-admission/zendesk-ai-agents)
- Agentforce and the Accusation an Agent Must Answer (/articles/conduct-admission/salesforce-agentforce-conduct)

TERMINOLOGY

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

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