

Copilot Coding Agents: Review Gates and Earned Exposure

A coding agent opens a pull request in a repository where nobody has seen its work before. The reviewer has no basis for calibration, and whatever the agent did well elsewhere does not follow it across the boundary. The common answer is a portable score: a rating, a trust tier, a vouch that one party hands to another. U.S. Provisional Application No. 64/117,812 takes a different route. In its Chapter 9, an agent earns admissibility rather than reputation, and it earns that admissibility by paying for it: counter-signing another agent's determination decrements the followed agent's own authorization budget, so the marginal cost of being followed rises with the rate of adoption. No score moves between agents at any point.

Standing that will not cross the repository line

An automated contributor opens a pull request against a codebase it has never touched. The diff looks plausible. The maintainer has no encounter history with this contributor and no prior corrections to weigh. Whatever the agent did well in another repository, under another owner, is unavailable at the moment it would be useful.

The usual repair is to make standing portable. Give the agent a score, a tier, or a vouch from a party the maintainer already trusts, and let that quantity travel with it. The trouble is what a portable quantity becomes once it exists. It gets compared against a

threshold, it gets summed, and it gets consumed as evidence by whatever stage will take a number, while the party that issued it spends nothing to issue it.

A second repair gates everything at the boundary instead: nothing merges without review and without passing checks. That works, and it is what serious engineering organizations do. A boundary gate is built to decide whether a particular change should land, evaluated per change against policy the repository owner controls.

Between those two answers sits a question neither is shaped to reach. When one agent elects to act under another agent's determination, who pays for that election, and what does the relying party actually acquire?

What the Copilot coding agent is built to do

GitHub Copilot, as publicly described, is an AI assistant for software development, and its agentic mode is designed to take a unit of work such as an issue and carry it through to a proposed change. As documented publicly, the coding agent works within a repository it has been granted access to and produces its output as a pull request, which enters the same path any other pull request enters: branch protection settings, required status checks, continuous integration, code owner rules, and human review before a merge.

State that design choice on its own terms rather than as a foil. The pull request is the right place to put an automated contributor, because it is already the repository's accountability surface. It is versioned and attributable, it carries a diff a human can read, and it can be blocked by policy the repository owner controls rather than the agent. An organization that already trusts its review gates gets to extend that trust to a new class of contributor without inventing new governance.

The category this serves is delegated execution under owner-controlled review, addressed to whether a given change should land. An earned-exposure architecture addresses a different question: what a relying party acquires, and what a relied-upon party spends, when one autonomous participant acts under another's judgment rather than its own.

Deference that is billed to the followed agent

Chapter 9 of the filed provisional describes a deference mechanism operating between semantic agents (100) with no rating, score, standing quantity, or credential passing between them. The setup is deliberately narrow. A first agent, the deferring agent, holds a request to dispatch an action and finds its own dispatch-authority predicate satisfied on recomputation from its memory field (102). It is already authorized to determine the matter for itself, and it nonetheless elects to execute under a determination recorded in a second agent's lineage field (104). An agent whose predicate fails, whose authorization gate (300) stands in the withheld state (310), or which has entered the non-executing cognitive mode (302) does not defer at all; the filing calls that a withholding rather than a following.

Two tests run before anything is given effect. The origin-disjointness test asks whether the two agents are assignable to a common origin-equivalence class (200); where they are, the record is appended as an attempted deference record and given no effect, decrementing nothing and appending nothing adverse to the followed agent. The retrievability-form verification then checks three ordered conjuncts against the presented followed determination record, among them a successor-continuity test on the attested epoch identifier and a lineage commitment carrying an extent value and an accumulator root over the ordered entries. That verification is of form and of retrievability and not of merits: the deferring agent does not retrieve the enumerated entries, does not re-derive the determination, and does not compare it against what its own admission evaluator (120) would have produced.

The inversion sits at counter-signature. When the followed agent elects to counter-sign, appending a deference counter-signature (902) as the second governed observation (610) of a matched pair (600), that append decrements its own authorization budget (404) under a deference decrement schedule declared in its signed policy object (112). The schedule declares an amount not less than that declared for issuance of a conduct evaluation artifact (116) by that agent, and is a monotonically non-decreasing function of the count of counter-signatures it has appended within a declared window, so that the marginal cost of being followed rises with the rate of adoption. The filing discloses two permitted forms for that function, one stepped and one linear in the count, each bounded above by a declared maximum and computed over the agent's own appends. The decrement occurs without regard to whether the deference was well founded; no party adjudicates that as a condition of it. The units decremented are the same units that gate the followed agent's own dispatch. Responsive to that budget satisfying a floor declared in its policy object, its authorization gate (300) is written to the withheld state (310) for an enumerated set of action classes and an escalation record is emitted to its principal. That budget is replenished by the ordinary procedure of the filing's Section 4.5 and by no other, so being followed spends a replenishable quantity.

What the deferring agent receives is not standing. Responsive to receipt of the counter-signature, and not otherwise, it retrieves a corrector-class enumeration (904): the origin-equivalence classes (200) recorded both as having corrected the followed agent in that action class and as having had that correction thereafter conformed to, membership requiring both parts of a two-part conformance filter (906). Each such class is written into the deferring agent's own replenishment register (402) as an admissible-but-unexercised entry, keyed on a class-defining reference rather than on any class identifier local to the followed agent. That entry has two effects and no third. An arriving conduct evaluation artifact (116) from an assignable asserting party (118) is admissible to the deferring agent's admission evaluator (120) notwithstanding no prior encounter history, and is not held in the pre-settlement inert state on that ground. For renewal, the entry is not counted while its exercise field stands at the unexercised value.

Where the two architectures diverge

The divergence is not about whether review should happen. Both positions treat unreviewed autonomous change as a bad idea.

The first divergence concerns the direction in which cost flows. The filing states its contrast directly: in conventional delegation and capability-attenuation architectures the cost falls upon the relying party and the party relied upon is protected, with attenuation narrowing only the capability passed downward. A review-gate model belongs to that family by design, and reasonably so: the maintainer accepts the review burden in exchange for policy the repository owner controls. The filed architecture places the metered consequence instead upon the party such an architecture protects, denominated in that party's own capacity to dispatch.

What each design causes to be acquired is the second divergence. A boundary gate reaches a decision about a change. The filed architecture writes something that persists about a relationship: the deferring agent acquires admissibility alone, written only upon a counter-signature that has been paid for, conferring nothing further until an artifact arrives and exercises it. On the exercised transition it appends an exercised transition record (908) binding the correction received back to the deference that earned the exposure through which it arrived, and where the class is absent from the replenishment register (402) for the declared window its authorization budget (404) is replenished under Section 4.5.

There is convergence too. Both treat the durable written record as the unit of account rather than a summary judgment about a participant, and both put policy in the hands of the party bearing the risk. The counter-signature window is declared in the deferring agent's signed policy object (112) and carried in the counter-signature window field (912) of the record as transmitted, so the bound agent reads the window from the record itself.

Coexisting, and what the filing does not address

The two fit together without contest. Repository review gates decide whether a change lands; the filed mechanism decides whether an agent may act under a peer agent's determination, and what that costs the peer. A team could leave every branch protection rule untouched while its agents settle among themselves which external evaluations they admit.

Name the limits. No adjudication of whether a deference was well founded is performed by either agent, by either principal, or by any other party as a condition of the decrement. Nor does the architecture yield a quantity that could be presented to a reviewer as a rating, because the conversion bar (502) forbids converting a deference abstention entry into a scalar value, a default value, or an operand of a threshold comparison, and forbids any admission or dispatch decision consuming a count of such records as a standing quantity. Where the followed agent declines or does not respond within the carried window, the deferring agent may still execute, appending an ungoverned reliance record and a reliance-bound narrowing record that narrows its own counterparty scope object by the categories declared in its policy; the filing states that narrowing is not a sanction and is released upon a counter-signature within the reinstatement window. Where two followed agents present determinations of differing outcome class on one conduct descriptor, the deferring agent appends a deference conflict record, adopts neither, and either determines the matter itself or withholds as a positive abstention. Quantities central to the mechanism, among them the counter-signature window, the observation interval, and the budget floor, are declared in the signed policy objects (112) of the participating agents rather than fixed by the filing, which leaves calibration to the deploying party.

Disclosure Scope

This article describes subject matter disclosed in U.S. Provisional Application No. 64/117,812, Chapter 9, covering deference and earned exposure between semantic agents. It is published for technical and defensive-publication purposes. It is not legal advice and it is not an offer to license.

References to the GitHub Copilot coding agent are to public materials and are used for comparison only; no relationship, endorsement, or infringement is asserted. Statements about that product are qualitative descriptions of its publicly documented purpose and category, not audits of its implementation, and should be checked against its own current documentation. The application referenced is pending, and no claim scope should be inferred from this description.

Deference and Earned Exposure ([/earned-exposure](#)) [All 49 steps → \(/inventive-steps\)](#)

Standing earned through recorded conduct; deference that survives across agents.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- [Earned Exposure and Deference Records: Standing Built From Recorded Conduct \(/articles/earned-exposure/earned-exposure-and-deference-records\)](#).

SECONDARY TECHNICAL

- [The Valuation-Free Ordinal Standing Quantity \(/articles/earned-exposure/valuation-free-ordinal-standing\)](#).
- [Determination-Invariance Under an Agent's Own Policy Succession \(/articles/earned-exposure/determination-invariance-policy-succession\)](#).
- [Elective Exposure Enrollment Without Deference \(/articles/earned-exposure/elective-exposure-without-deference\)](#).

- [Interventional Attribution by Receipt-Path Control \(/articles/earned-exposure/interventional-attribution-receipt-path\)](/articles/earned-exposure/interventional-attribution-receipt-path).
- [The Portable Exoneration Record \(/articles/earned-exposure/portable-exoneration-record\)](/articles/earned-exposure/portable-exoneration-record).
- [Attribution, Exoneration, and the Domain Baseline: An Ordered Branch Over the Latency Differential \(/articles/earned-exposure/attribution-versus-exoneration-branch-domain-baseline\)](/articles/earned-exposure/attribution-versus-exoneration-branch-domain-baseline)
- [Bounded Count of Dispatches Per Deference Record: Capping How Many Executions One Counter-Signed Adopted Determination Authorizes \(/articles/earned-exposure/bounded-count-dispatches-deference-record\)](/articles/earned-exposure/bounded-count-dispatches-deference-record)
- [Coarse Lineage-Segment Granularity: Verifying a Tested-Entry Enumeration Without Disclosing Which Lineage Entries Were Tested \(/articles/earned-exposure/coarse-lineage-segment-granularity-tested-entry\)](/articles/earned-exposure/coarse-lineage-segment-granularity-tested-entry)
- [Correction-Exposure Decay: An Authorization Quantity That Depletes With Agent Operation and Renews Only on Recorded External Evaluation \(/articles/earned-exposure/correction-exposure-decay-authorization-quantity\)](/articles/earned-exposure/correction-exposure-decay-authorization-quantity)
- [Cause-Keyed Gate Restoration: A Distinct Return Path for Decay-Induced Withholding of Agent Authorization \(/articles/earned-exposure/distinct-return-path-decay-induced-withholding\)](/articles/earned-exposure/distinct-return-path-decay-induced-withholding)
- [Edge Severance Events and the Three-Valued Survival State: Resolving Whether a Counterparty Relationship Outlived a Payment-Class or Obligation-Class Termination \(/articles/earned-exposure/edge-severance-events-three-valued-survival\)](/articles/earned-exposure/edge-severance-events-three-valued-survival)
- [Emitting the Deference Correction Linkage Record to the Followed Agent: Reporting a Downstream Correction Without Producing a Rating \(/articles/earned-exposure/emission-deference-correction-linkage-record-followed\)](/articles/earned-exposure/emission-deference-correction-linkage-record-followed)
- [The No-Address Omission Branch: What a Deferring Agent Records When an Exposure Dispatch Reaches Nobody \(/articles/earned-exposure/exposure-dispatch-no-address-omission-branch\)](/articles/earned-exposure/exposure-dispatch-no-address-omission-branch)
- [Channel-Baseline Normalization: Taking Channel Overhead Out of the Latency Differential \(/articles/earned-exposure/latency-differential-channel-baseline-normalization\)](/articles/earned-exposure/latency-differential-channel-baseline-normalization)
- [Reason-Type Resolution From Severance-Survival States: Typing an Agent's Relationship Edges Without Asking the Counterparty \(/articles/earned-exposure/reason-type-resolution-survival-states\)](/articles/earned-exposure/reason-type-resolution-survival-states)
- [Reason-Type Retyping by a Superseding-Type Entry: Recomputing an Edge's Reason-Type on Later Severance Evidence Without Disturbing Modifications Already Applied \(/articles/earned-exposure/reason-type-retyping-superseding-type-entry\)](/articles/earned-exposure/reason-type-retyping-superseding-type-entry)
- [Renewal-Side Replenishment of the Authorization Budget: Restoring an Agent's Metered Spending Capacity Only on Exposure to an Uncounted Origin-Equivalence Class \(/articles/earned-exposure/renewal-side-replenishment-authorization-budget\)](/articles/earned-exposure/renewal-side-replenishment-authorization-budget)
- [Following Does Not Count as Being Tested: Retention of the Short Decay Constant When an Agent Defers, and the Action Class That Stays Untested \(/articles/earned-exposure/retention-short-decay-constant-deference-class\)](/articles/earned-exposure/retention-short-decay-constant-deference-class)

- [Standing-Determination Designation: Substituting a Verified Counterparty Determination for One Conjunct of the Dispatch-Authority Predicate \(/articles/earned-exposure/standing-determination-designation-substitution-inside-dispatch\)](/articles/earned-exposure/standing-determination-designation-substitution-inside-dispatch).
- [Uncounted-Class Renewal: Gating Agent Authorization on a Windowed Register of Origin-Equivalence Classes \(/articles/earned-exposure/uncounted-class-renewal-gated-windowed-register\)](/articles/earned-exposure/uncounted-class-renewal-gated-windowed-register).

APPLICATIONS • GENERAL

- [Vendor Qualification Without Ratings: Standing Earned From Recorded Conduct \(/articles/earned-exposure/vendor-qualification-without-ratings\)](/articles/earned-exposure/vendor-qualification-without-ratings).
- [Credentialing Autonomous Practitioners: Deference That Survives Across Agents \(/articles/earned-exposure/professional-credentialing-agents\)](/articles/earned-exposure/professional-credentialing-agents).

APPLICATIONS • SPECIFIC

- [Marketplace Seller Reputation Systems: Presented Scores Versus Earned Admissibility \(/articles/earned-exposure/marketplace-seller-reputation\)](/articles/earned-exposure/marketplace-seller-reputation).
- [Letta: Agent Memory and Standing Earned From Conduct \(/articles/earned-exposure/letta-agent-memory\)](/articles/earned-exposure/letta-agent-memory).
- [Trustpilot Signals and Conduct-Derived Admissibility \(/articles/earned-exposure/trustpilot-signals\)](/articles/earned-exposure/trustpilot-signals).
- [Copilot Coding Agents: Review Gates and Earned Exposure \(/articles/earned-exposure/github-copilot-agent\)](/articles/earned-exposure/github-copilot-agent).
- [Supplier Risk Scores and Recorded Counterparty Conduct \(/articles/earned-exposure/supplier-risk-scoring\)](/articles/earned-exposure/supplier-risk-scoring).
- [Verified Profiles and Standing Built From Recorded Conduct \(/articles/earned-exposure/linkedin-verification\)](/articles/earned-exposure/linkedin-verification).

TERMINOLOGY

- [counterparty identity record \(/glossary#counterparty-identity-record\)](/glossary#counterparty-identity-record).
- [deference record \(/glossary#deference-record\)](/glossary#deference-record).
- [edge \(/glossary#edge\)](/glossary#edge).
- [reason-type \(/glossary#reason-type\)](/glossary#reason-type).
- [severance event \(/glossary#severance-event\)](/glossary#severance-event).
- [standing quantity \(/glossary#standing-quantity\)](/glossary#standing-quantity).
- [untested class \(/glossary#untested-class\)](/glossary#untested-class).

Deference and Earned Exposure overview → (/earned-exposure).