

Matched-Pair-Gated Corpus Contribution

AI agents improve by retraining on material other parties contribute, yet a contributor who later severs the relationship or is quarantined leaves no reliable trace of which models absorbed its data or on what terms. Matched-Pair-Gated Corpus Contribution, disclosed in U.S. Provisional Application No. 64/117,812, closes that gap. A receiving agent admits a counterparty's training-corpus contribution only as the second observation of a bilateral matched pair whose first observation is a contribution-scope offer the agent itself emits, naming the corpus policy version, the scope partitions, and the action classes that will govern any retrained model. Recognition yields a settled contribution identifier, a digest over the settlement-lineage entry, and no counterparty-sourced material enters a corpus without one. Every retraining event then records exactly which settled identifiers it incorporated.

1. The Gap

An autonomous agent that retrains or adapts its own inference models does not learn in a vacuum. It ingests material that other parties supply, and that ingestion opens two liabilities most pipelines never record at the moment they arise.

One is downstream. When a contributor later severs the relationship or is placed under quarantine, the agent has to know precisely which of its models absorbed that contributor's material, and on what terms the material was admitted. Provenance and lineage tooling can describe where data came from, but it usually does so as after-the-fact metadata assembled in a central store. It rarely produces a record that both parties fixed at the instant of contribution, and it rarely refuses ingestion when no such record exists.

The other liability is symmetric. A contributor can later deny having supplied anything, or an agent can claim material was authorized when it was not. Without a mutually produced, non-repudiable artifact fixing the terms of contribution, that dispute cannot be resolved from either party's records alone.

The mechanism disclosed here treats a corpus contribution as a settlement rather than a file transfer. Counterparty-sourced material does not cross into a corpus unless it arrives bound to a bilateral, cryptographically recognized record that names the policy version, the scope partitions, and the action classes under which any resulting model will operate. That record is the admission ticket. The same record later becomes the key by which the agent traces, and if necessary encumbers, every endpoint that consumed the contribution.

2. Mechanism

The receiving agent (100) admits a counterparty's training-corpus contribution only as the second governed observation (610) of a matched pair (600). Present the same material in any other form and it reaches no corpus and enters no retraining event, because no settled identifier ever attaches to it.

The first governed observation (608) is a contribution-scope offer, and the receiving agent is the party that emits it. The agent that will carry the material into its own models is therefore the one that states the terms. The offer enumerates three things at

once: the version identifier of the corpus policy in force, the scope partitions in which a model retrained on the contribution will be held and registered, and the action classes for which that model will be dispatched. This is not an open-ended request for data. It is a bounded proposal that fixes, before anything is contributed, the governance envelope any retrained model will occupy.

The counterparty replies as the second governed observation (610). Its payload carries a digest of the contributed material and a reference to the offer it answers. The counterparty is not pouring data into an unbounded pool; it is assenting, on the record and under its own authority credential, to the exact policy version, partitions, and action classes the offer named.

Recognition of the pair follows the same rule that governs every matched-pair settlement in the filing. On recognition, each party appends a settlement-lineage entry, held independently on each side, and that entry is the authority for the contribution. No platform and no third party takes part in producing it.

From that entry the receiving agent computes a settled contribution identifier as a digest over it, and against the identifier it records the contributor identity, the material digests, the policy version, and the scopes. Because the identifier is a digest of the settlement, a settlement differing in any recorded element produces a different identifier, and the single value carries who contributed, what was contributed, under which policy, and for which scopes.

The gate follows: no counterparty-sourced material is admitted for which no settled identifier exists. This is an admissibility rule at the corpus boundary, not a logging convention, and it is expressly bounded. It does not touch user-authored, user-curated, or user-designated material, nor the other retraining triggers disclosed elsewhere in the applicant's tool-registry disclosure. What it constrains is material sourced from a counterparty, the one category that carries the two liabilities above.

The record does not end at admission. Each retraining event records the set of settled identifiers it incorporated, which is its recorded contribution set, and the correspondence between an identifier and the corpus material it stands for is fixed by digest equality alone. If a piece of corpus material has a digest recorded against a settled identifier, that identifier joins the retraining event's contribution set. If its digest matches none, no settled identifier exists for it and the gate holds.

What emerges is a traceable line from a bilateral settlement, through a content-addressed identifier, into the recorded contribution set of a specific retraining event, and on to the endpoint that descends from it. Because the identifier binds contributor identity and agreed scopes at the moment of settlement, the agent can later answer from its own records alone which endpoints hold a given counterparty's material and under what terms it was admitted.

3. Operating Parameters

The mechanism declares structural parameters, not numeric thresholds, and the filing states them exactly.

The contribution-scope offer carries three enumerated fields: the corpus policy version identifier, the scope-partition identifiers, and the action-class identifiers. Those three are the terms the counterparty's second observation assents to.

The settled contribution identifier is a digest over the settlement-lineage entry. Contributed corpus material is itself identified by a digest carried in the second observation, and matching between corpus material and a settled identifier is by digest equality alone. Recording and matching turn on one digest function: the function that produced the recorded digests is the function under which later material is tested.

Recognition of the pair inherits the settlement parameters the filing applies to every matched pair. Two of them do the work here: a temporal proximity window within which the first and second observations must pair, and, where the two agents are not physically co-located, a scope-partition correspondence that stands in for spatial coincidence. A contribution whose pairing is not recognized, whether it is refused or falls outside the window, produces no settlement-lineage entry. With no entry there is no settled contribution identifier, and with no identifier nothing is admitted.

4. Composition

This mechanism lives in the tool-provenance material of the filing, and it shares a substrate with the mechanisms around it: the endpoint record the agent maintains for each registered inference artifact.

Its closest companion is the contribution-encumbrance machinery in the same section. On severing a counterparty or writing a quarantine attribute, the receiving agent enumerates every endpoint whose recorded contribution set contains a settled contribution identifier of that counterparty and writes a contribution-encumbered attribute into each, narrowing that counterparty's scope so that dispatch to the encumbered endpoint becomes inadmissible for any action whose affected party maps to the counterparty. That enumeration is possible only because the gate guarantees that every counterparty ingestion left a settled identifier tied to a contributor and a set of scopes. The agent deletes nothing and retrains nothing on severance. It narrows access to what it already holds, and it can do so because the settlement fixed the association in advance.

Above the gate sits the matched-pair settlement framework the filing uses for all bilateral exchange. A corpus contribution settles through the identical governed-observation and settlement-lineage machinery that settles other transactions, so the same recognition and non-repudiation properties carry over with no separate protocol. The positive-abstention discipline applies too: when the narrowing withholds a

dispatch to an encumbered endpoint, the withholding is appended as a positive abstention rather than a denial. And the scope-partitioned architecture closes the loop, because the partitions named in the offer are the same partitions in which the retrained model is held and registered.

A single settlement event, then, propagates its terms in three directions at once: into corpus admission, into the identity of every retraining event that draws on the contribution, and into the later governance of every endpoint that descends from it.

5. Prior-Art Distinction

Four established bodies of work sit near this mechanism, and none supplies what it does.

Bilateral value settlement, including payment channels and escrow protocols, does settle obligations between two parties, but the settled object is value or state, and settlement generally leans on an intermediary, a consensus authority, or pre-negotiated session state. Here the settled object is the admission of training material under a named scope envelope, and recognition of a matched pair is the whole of the settlement, leaning on none of the three.

Data provenance and lineage tooling, from dataset cards to lineage graphs to content-addressed dataset registries, records where training data originated. It typically records that as descriptive metadata assembled after ingestion, often centrally. It does not gate ingestion on a mutually signed settlement, does not bind a contribution to a policy version and an action-class scope enforced at retraining, and does not, when a contributor later withdraws, mechanically identify the affected models.

Data-licensing and consent frameworks fix terms between a data provider and a consumer, but a license or a consent form is a legal instrument that sits outside the pipeline. Nothing in the runtime refuses to ingest material for which no machine-

checkable settlement exists, and revoking consent does not by itself resolve which models incorporated the licensed material.

Federated learning and contribution-attribution schemes track per-participant contributions to a shared model, yet they usually run under a central coordinator or aggregator that assigns and reconciles the updates. This mechanism is peer-to-peer with no coordinator, and the identifier it produces keys the later governance and encumbrance of a contribution rather than a reward or a weighting of it.

These are real and useful lines of work. The distinction is structural: only this mechanism refuses corpus admission at the boundary unless a bilateral, cryptographically recognized settlement has produced a content-addressed identifier binding contributor, material, policy version, and scope, and then carries that identifier into the identity of every retraining event by digest equality alone. Nothing here should be read as asserting that any product or party practices, or infringes, the disclosed subject matter.

6. Disclosure Scope

This mechanism is disclosed in U.S. Provisional Application No. 64/117,812, Section 10.5, paragraph [0403], under the heading "Matched-pair-gated corpus contribution and the settled contribution identifier," together with the surrounding tool-provenance and contribution-encumbrance paragraphs of that section.

What is disclosed: admission of a counterparty's training-corpus contribution only as the second governed observation of a matched pair; the contribution-scope offer as the first observation, enumerating the corpus policy version, the scope partitions in which a retrained model will be held and registered, and the action classes for which it will be dispatched; the counterparty's observation carrying a digest of the contributed material and a reference to the offer; the settlement-lineage entry each party appends on recognition; the settled contribution identifier computed as a digest over that entry and

recording the contributor identity, the material digests, the policy version, and the scopes; the admissibility rule barring counterparty-sourced material for which no settled identifier exists; and each retraining event's recorded contribution set, matched to corpus material by digest equality alone.

What is disclaimed here: the rule does not limit the ingestion of user-authored, user-curated, or user-designated material, or the other retraining triggers disclosed in the applicant's tool-registry disclosure. This article is a defensive technical publication describing the filed disclosure. It does not enlarge that disclosure, and specific sub-embodiments developed in the applicant's fuller working papers but absent from the filed provisional are neither described nor published here, being reserved to the applicant.

Governed Matched-Pair Settlement

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

(/governed-settlement)

Two agents settle on matched signed observations, by continuity, with no clearer.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- [Matched-Pair Settlement Between Agents: Bilateral Closure Without a Clearer \(/articles/governed-settlement/matched-pair-settlement-without-a-clearer\)](#).

SECONDARY TECHNICAL

- [The Derived-Artifact Provenance Chain and Chain-Head Identifier \(/articles/governed-settlement/derived-artifact-provenance-chain\)](#).
- [Provenance Withdrawal and the Provenance-Withdrawn Attribute \(/articles/governed-settlement/provenance-withdrawal\)](#).
- [Matched-Pair-Gated Corpus Contribution \(/articles/governed-settlement/matched-pair-gated-corpus-contribution\)](#).

- [Partition-Scoped Escrow Release Gated on Intersection Non-Emptiness \(/articles/governed-settlement/partition-scoped-escrow-release\)](/articles/governed-settlement/partition-scoped-escrow-release).
- [The First-Encounter Settled Pair and Promotion to the Persistent Tier \(/articles/governed-settlement/first-encounter-settled-pair\)](/articles/governed-settlement/first-encounter-settled-pair).

TERMINOLOGY

- [governed observation \(/glossary#governed-observation\)](/glossary#governed-observation).
- [matched pair \(/glossary#matched-pair\)](/glossary#matched-pair).
- [scope partition \(/glossary#scope-partition\)](/glossary#scope-partition).

[Governed Matched-Pair Settlement overview → \(/governed-settlement\)](/governed-settlement)