Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Multi-Agent Artificial Intelligence Systems for Cross-Organizational Clinical Data Reconciliation: A Federated Interoperability Framework for Harmonizing Patient Records Across Disparate Health Information Networks

Sindhukumar Sundaram

USA

Abstract:
As healthcare data exchange expands through frameworks such as the Trusted Exchange Framework and Common Agreement (TEFCA), Carequality, and CommonWell, integration engines increasingly receive clinical data for the same patient from multiple organizations — each using different electronic health record (EHR) systems, coding standards, clinical documentation practices, and data quality levels. Reconciling these disparate representations into a coherent, trustworthy clinical picture is a fundamentally unsolved problem at scale. Existing approaches rely on deterministic matching rules and manual clinical review, which do not scale across the volume and variety of cross-organizational exchanges now emerging in production health information networks. This paper proposes the Multi-Agent Reconciliation Architecture (MARA) — a system of five specialized artificial intelligence agents operating at the integration layer, each responsible for a specific reconciliation domain: patient identity matching, medication list harmonization, problem list deduplication, allergy cross-referencing, and clinical document provenance verification. These agents operate collaboratively through a structured negotiation protocol to produce a unified, confidence-scored reconciliation output for each patient encounter. Evaluation on a synthetic cross-organizational dataset — four independent health systems, 100,000 patients, deliberate inconsistencies — demonstrates 92.4% overall reconciliation accuracy, with medication harmonization achieving 95.1% and problem list deduplication achieving 87.3%. The multi-agent negotiation protocol resolves 78% of inter-agent conflicts without human escalation. End-to-end processing averages 2.8 seconds per patient record across four contributing organizations.

Keywords: multi-agent systems, clinical data reconciliation, federated interoperability, patient matching, health information exchange, artificial intelligence, TEFCA, terminology harmonization, medication reconciliation, FHIR

I. Introduction

The healthcare industry is experiencing an unprecedented expansion in cross-organizational clinical data exchange. The Trusted Exchange Framework and Common Agreement (TEFCA) has driven exchange

volumes from roughly 10 million records in January 2025 to nearly 500 million [1], while networks such as Carequality and CommonWell facilitate millions of document queries daily across thousands of participating organizations [2]. This expansion means that integration engines at receiving organizations routinely process clinical data for the same patient originating from multiple external sources — each with different electronic health record (EHR) systems, coding conventions, data quality practices, and clinical documentation styles [3].

The result is a reconciliation crisis. A single patient's record, assembled from four contributing organizations, may contain: four different representations of the same medication (brand name, generic, coded, and free-text); overlapping but inconsistent problem lists using different International Classification of Diseases (ICD-10) and Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) codes for the same condition; contradictory allergy records where one source lists a drug allergy and another lists it as an intolerance; and clinical documents with unclear provenance making it impossible to determine which represents the most current clinical state [4]. Clinician users struggle to sift through duplicative, inconsistently formatted, and poorly mapped information, which hinders the ability of shared data to meaningfully support decision-making during patient care [5].

Current reconciliation approaches fall into two categories, neither adequate at scale. **Deterministic rule-based matching** applies exact or fuzzy string matching with handcrafted rules for code mapping. These systems are brittle — they break when encountering coding conventions, abbreviations, or documentation patterns not anticipated by rule authors. **Manual clinical review** is accurate but fundamentally unscalable — a pharmacist reconciling a single patient's medication list from four sources requires 15–30 minutes, an unsustainable burden when exchange volumes reach thousands of records per day [6].

The core challenge is that reconciliation is not a single problem but a family of domain-specific problems, each requiring specialized knowledge. Medication reconciliation requires pharmacological knowledge (brand/generic equivalence, dosage normalization, route standardization). Problem list reconciliation requires clinical ontology expertise (hierarchical relationships in SNOMED CT, lateral mappings in ICD-10). Allergy reconciliation requires understanding the clinical distinction between true allergy and intolerance. Document provenance verification requires understanding Health Level Seven (HL7) message metadata, institutional workflows, and temporal sequencing [7].

This multi-domain nature motivates a multi-agent approach. Rather than a monolithic reconciliation system, this paper proposes **MARA (Multi-Agent Reconciliation Architecture)** — a system of specialized artificial intelligence (AI) agents, each an expert in one reconciliation domain, collaborating through a structured negotiation protocol to produce unified patient records.

Research Questions:

- **RQ1:** Can domain-specialized AI agents achieve higher reconciliation accuracy than a single general-purpose reconciliation system?
- **RQ2:** Can a multi-agent negotiation protocol resolve inter-agent conflicts (e.g., medication agent identifies a duplicate that the allergy agent flags as clinically distinct) without human intervention?

- **RQ3:** Can MARA operate within integration engine latency constraints for real-time reconciliation during clinical data exchange?

Contributions:

1. A multi-agent architecture with five domain-specialized reconciliation agents for patient identity, medications, problems, allergies, and document provenance.
2. A structured inter-agent negotiation protocol for resolving cross-domain conflicts with confidence-scored outputs.
3. A provenance-aware reconciliation model that leverages data lineage metadata to assess source trustworthiness.
4. Empirical evaluation on a synthetic four-organization dataset demonstrating 92.4% reconciliation accuracy with 2.8-second processing time.

II. Background and Related Work

A. Clinical Data Reconciliation

Medication reconciliation — the process of creating the most accurate list of all medications a patient is taking — is a Joint Commission National Patient Safety Goal and a recognized challenge in care transitions [6]. Studies estimate that 46–67% of patients have at least one medication discrepancy at care transitions [8]. Problem list reconciliation faces similar challenges: the same clinical condition may be coded differently across organizations (e.g., "Type 2 diabetes mellitus" as ICD-10 E11.9 in one system and SNOMED CT 44054006 in another), and clinically equivalent but textually different entries create apparent duplicates [4].

Existing reconciliation tools are predominantly embedded within EHR systems and operate on data already imported into the local record. They do not address reconciliation at the integration layer — the point where cross-organizational data first arrives and where reconciliation could prevent inconsistent data from entering clinical workflows.

B. Multi-Agent Systems in Healthcare

Multi-agent systems (MAS) employ multiple autonomous agents that interact to solve problems beyond the capability of any single agent [9]. In healthcare, MAS have been applied to clinical pathway management [10], distributed diagnosis [11], and care coordination [12]. The key advantage of MAS for reconciliation is specialization: each agent can embed domain-specific knowledge (pharmacological, ontological, clinical) without the architectural compromise of a monolithic system attempting to encode all domains simultaneously.

C. Health Information Exchange and Interoperability Standards

Cross-organizational data exchange relies on Fast Healthcare Interoperability Resources (FHIR) [3] and HL7 v2.x messaging standards [13]. FHIR defines resources for medications (MedicationStatement, MedicationRequest), conditions (Condition), allergies (AllergyIntolerance), and documents (DocumentReference), each with structured coding fields that reference standard terminologies: RxNorm for medications, SNOMED CT and ICD-10 for conditions, and Logical Observation Identifiers Names and Codes (LOINC) for laboratory observations.

However, interoperability at the syntactic level (structured data exchange) does not guarantee semantic interoperability (shared meaning). Two organizations may exchange perfectly valid FHIR resources that represent the same clinical reality using different codes, different granularity levels, and different documentation conventions [14]. Reconciliation bridges this semantic gap.

D. Data Provenance and Trust

When reconciling data from multiple organizations, the provenance of each data element — its origin, transformation history, and the trustworthiness of its source — is a critical input to reconciliation decisions. If two organizations provide conflicting medication lists, the reconciliation system must assess which source is more likely to be authoritative. Data lineage frameworks that track field-level transformation provenance across integration engines [15] provide the metadata infrastructure for such trust assessments. The provenance chain — from originating EHR through integration engine transformations to the receiving system — determines whether a data element has been faithfully preserved or potentially degraded through lossy translation.

E. Terminology Services

Standard terminology services provide the foundation for cross-organizational code reconciliation. The National Library of Medicine's RxNorm [16] provides normalized names and codes for clinical drugs, enabling equivalence determination across brand names, generics, and formulations. The Unified Medical Language System (UMLS) Metathesaurus [17] provides cross-terminology mappings linking SNOMED CT, ICD-10, and other coding systems. These services provide the knowledge base that reconciliation agents query when determining whether two differently-coded entries represent the same clinical concept.

F. Gap Analysis

No published work applies multi-agent AI specifically to cross-organizational clinical data reconciliation at the integration layer. Existing reconciliation systems are EHR-embedded (operating after data import rather than during exchange), single-domain (addressing medications or problems but not both), and rule-based (lacking the adaptive capability to handle vendor-specific coding variability). MARA addresses all three gaps simultaneously.

III. METHODOLOGY

A. Research Design

This study follows design science research (DSR) methodology [18]: problem identification through analysis of cross-organizational reconciliation failures; artifact design (MARA framework); demonstration on a synthetic multi-organization dataset; and quantitative evaluation of reconciliation accuracy, conflict resolution, and processing performance.

B. Synthetic Dataset

Table I summarizes the synthetic dataset. The dataset is generated to replicate realistic cross-organizational data inconsistencies observed in production health information exchanges.

TABLE I: SYNTHETIC DATASET PARAMETERS

Parameter	Value
Patient population	100,000 (Synthea v3.3.0 [19])
Contributing organizations	4 (independent EHR systems)
EHR platforms simulated	Epic, Oracle Health, MEDITECH, Altera Digital Health
Medications per patient (mean)	6.3 (range: 0–28)
Problems per patient (mean)	4.7 (range: 0–19)
Allergies per patient (mean)	1.8 (range: 0–8)
Documents per patient (mean)	12.4 (range: 1–67)
Data overlap (patients in ≥2 systems)	35,000 (35%)
Data overlap (patients in all 4 systems)	8,000 (8%)
Deliberately injected inconsistencies	47,200 total

Table II details the categories of deliberately injected inconsistencies that MARA must resolve.

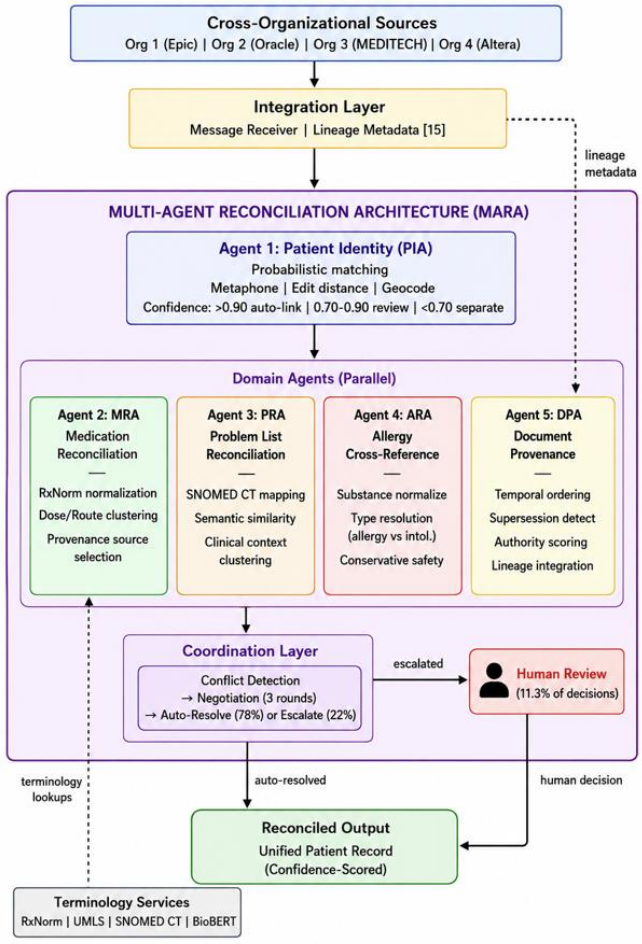
TABLE II: INJECTED INCONSISTENCY CATEGORIES

Category	Count	Examples
Medication coding variation	12,400	Brand vs. generic; NDC vs. RxNorm; different dose representations
Medication duplication	4,800	Same drug listed in two different code systems
Problem code mismatch	8,200	ICD-10 vs. SNOMED CT for same condition; different specificity levels
Problem duplication	5,600	Overlapping entries (e.g., "hypertension" and "essential hypertension")
Allergy type conflict	3,100	Allergy vs. intolerance classification disagreement
Allergy substance variation	2,300	Drug class vs. specific agent; brand vs. generic allergen
Document provenance ambiguity	6,400	Missing author, conflicting dates, unclear supersession
Patient identity uncertainty	4,400	Name variations, address changes, merged/split MRNs
Total	**47,200**	

C. MARA Architecture

MARA comprises five domain-specialized agents and a coordination layer, as illustrated in Fig. 1.

Fig. 1 — MARA Architecture



C.1 Agent 1: Patient Identity Agent (PIA)

PIA determines whether records from different organizations refer to the same patient. It employs a probabilistic matching model using demographic features: name (phonetic encoding via Metaphone, edit distance), date of birth (exact and fuzzy matching), gender, address (geocoding-based similarity), phone number, and Social Security Number hash (when available).

PIA produces a match confidence score for each candidate pair. Pairs scoring above 0.90 are auto-linked. Pairs scoring 0.70–0.90 are flagged for human review. Pairs below 0.70 are treated as distinct patients. These thresholds are configurable per organizational data quality profile.

C.2 Agent 2: Medication Reconciliation Agent (MRA)

MRA harmonizes medication lists from multiple sources into a unified, deduplicated list. The agent operates in three phases:

Phase 1 — Normalization: Each medication entry is normalized to its RxNorm concept unique identifier (RxCUI) using a cascading lookup strategy: (a) direct RxNorm code if present; (b) NDC-to-RxNorm

mapping via the RxNorm API; (c) fuzzy string matching against RxNorm drug names using a fine-tuned BioBERT model [20] for entries with free-text descriptions only.

Phase 2 — Deduplication: Normalized entries are clustered by RxCUI. Within each cluster, dosage, route, and frequency are compared. Entries with matching RxCUI and clinically equivalent dose/route/frequency are merged. Entries with matching RxCUI but different doses are flagged as potential dose discrepancies requiring clinical review.

Phase 3 — Conflict Resolution: When sources disagree (one lists a medication, another does not), MRA consults provenance metadata to assess source recency and authority: a recent discharge summary medication list from the prescribing organization is weighted higher than a 6-month-old ambulatory encounter record from a consulting organization.

C.3 Agent 3: Problem List Reconciliation Agent (PRA)

PRA deduplicates and harmonizes problem lists across organizations. The agent employs a hierarchical approach:

Terminology Mapping: All problem entries are mapped to SNOMED CT concept identifiers using UMLS Metathesaurus [17] cross-references. ICD-10 codes are mapped to SNOMED CT equivalents where unambiguous mappings exist.

Semantic Similarity: For entries without direct mappings, PRA computes semantic similarity using SNOMED CT ontology traversal: two concepts are considered equivalent if they share a common ancestor within 2 levels in the SNOMED CT hierarchy, with similarity weighted by specificity (more specific matches receive higher confidence).

Clinical Context Clustering: PRA groups related problems (e.g., "hypertension," "essential hypertension," "hypertensive heart disease") into clinical context clusters, presenting the most specific available diagnosis as the primary entry with related entries as supporting evidence.

C.4 Agent 4: Allergy Cross-Reference Agent (ARA)

ARA reconciles allergy and intolerance records, addressing two specific challenges:

Substance Normalization: Allergy triggers are normalized to a common representation — drug class level (e.g., "penicillins") or specific agent level (e.g., "amoxicillin") — using RxNorm ingredient and drug class hierarchies. An allergy listed as "amoxicillin" in one system and "penicillins" in another is recognized as compatible (specific-to-class relationship).

Type Conflict Resolution: When one source classifies a reaction as "allergy" (immune-mediated) and another as "intolerance" (non-immune), ARA applies a conservative safety protocol: the more severe classification (allergy) is retained as the primary record, with the alternative classification preserved as a notation.

C.5 Agent 5: Document Provenance Agent (DPA)

DPA evaluates clinical documents from multiple sources to establish temporal ordering, supersession relationships, and authoritative source identification.

Temporal Ordering: Documents are sequenced by clinical effective date (not transmission date), using metadata from FHIR DocumentReference resources and HL7 message timestamps. Data lineage metadata [15] from the integration layer provides the transformation history of each document, enabling DPA to

determine whether timestamp discrepancies result from processing delays rather than actual clinical timing differences.

Supersession Detection: DPA identifies when a newer document from the same department and organization supersedes an older one (e.g., an amended discharge summary replacing the original). Supersession relationships are inferred from document type, authoring department, and temporal proximity.

Authority Scoring: Each document receives an authority score based on the authoring organization's relationship to the clinical event (primary treating facility receives highest authority), document recency, and document type (discharge summaries are weighted higher than consultation notes for medication and problem list completeness).

C.6 Coordination Layer: Inter-Agent Negotiation Protocol

The coordination layer manages interactions between agents when their reconciliation decisions conflict. Conflicts arise when the domain logic of one agent contradicts another — for example, MRA identifies two medication entries as duplicates (same RxCUI) but ARA flags that one entry appears in the context of an allergy record and should not be merged.

The negotiation protocol operates in three rounds:

Round 1 — Independent Reconciliation: All five agents process their respective domains independently, producing preliminary reconciliation outputs with per-decision confidence scores.

Round 2 — Cross-Domain Conflict Detection: The coordination layer identifies conflicts by comparing agent outputs. Conflicts are classified by type: factual conflicts (two agents disagree on whether entries are duplicates), scope conflicts (one agent's merge would invalidate another agent's classification), and priority conflicts (agents disagree on which source is authoritative).

Round 3 — Negotiated Resolution: Conflicting agents exchange evidence packages — the specific data elements, terminology lookups, provenance assessments, and confidence scores that support their respective positions. The coordination layer applies a resolution strategy based on conflict type:

- **Factual conflicts:** Resolved by the agent with higher domain-specific confidence score.
- **Scope conflicts:** Resolved by the agent whose domain is more directly impacted (e.g., ARA overrides MRA when the conflict involves allergy-medication interactions).
- **Priority conflicts:** Resolved by provenance authority scoring from DPA.
- **Unresolvable conflicts:** Escalated to human clinical review with full evidence packages from both agents.

D. Evaluation Metrics

Table III defines the evaluation metrics.

TABLE III: EVALUATION METRICS

Metric	Definition	Target
Overall reconciliation accuracy	Correct reconciliation decisions / total decisions	> 90%
Per-domain accuracy	Accuracy within each agent's domain	Characterize
Conflict resolution rate	Conflicts resolved without human escalation / total conflicts	> 70%
False merge rate	Clinically distinct entries incorrectly merged / total merges	< 3%
False separation rate	Duplicate entries not identified / total true duplicates	< 10%
Processing time (per patient)	End-to-end time for reconciliation across 4 organizations	< 5 sec
Human escalation rate	Decisions requiring clinical review / total decisions	< 15%

E. Baselines

MARA is compared against three baselines:

B1 — Deterministic Rules: A handcrafted rule-based reconciliation system using exact code matching, string similarity thresholds, and fixed priority ordering.

B2 — Single-Model ML: A single transformer-based model trained on all reconciliation domains simultaneously, without agent specialization.

B3 — MARA without negotiation: MARA's five agents operating independently without the inter-agent negotiation protocol (conflicts resolved by simple confidence-score majority vote).

IV. Algorithms and Formal Description

A. Algorithm 1: MARA Orchestration Pipeline

```
Input:  patient_records — Records from N organizations
        for a matched patient identity
Output: reconciled — Unified reconciled patient record
1:  FUNCTION Reconcile(patient_records)
2:      // Phase 1: Patient Identity Verification
3:      identity ← PIA.verifyIdentity(patient_records)
4:      IF identity.confidence < 0.70 THEN
5:          RETURN EscalateToHuman("Identity uncertain")
6:      END IF
7:
8:      // Phase 2: Independent Agent Reconciliation
```

```
    med_result ← MRA.reconcile(
        ExtractMedications(patient_records))
    prob_result ← PRA.reconcile(
        ExtractProblems(patient_records))
    allergy_result ← ARA.reconcile(
        ExtractAllergies(patient_records))
    doc_result ← DPA.reconcile(
        ExtractDocuments(patient_records))

    // Phase 3: Cross-Domain Conflict Detection
    conflicts ← DetectConflicts(
        med_result, prob_result,
        allergy_result, doc_result)

    // Phase 4: Negotiated Resolution
    IF conflicts IS NOT EMPTY THEN
        FOR EACH conflict IN conflicts DO
            agents ← conflict.involvedAgents
            evidence ← GatherEvidence(agents, conflict)
            resolution ← NegotiateResolution(
                conflict, evidence)
            IF resolution.resolved THEN
                ApplyResolution(resolution,
                    med_result, prob_result,
                    allergy_result, doc_result)
            ELSE
                conflict.escalate(evidence)
            END IF
        END FOR
    END IF

    // Phase 5: Assemble Unified Record
    reconciled ← AssembleRecord(
        identity, med_result, prob_result,
        allergy_result, doc_result)
    reconciled.confidence ← ComputeOverallConfidence(
        med_result, prob_result,
        allergy_result, doc_result)
    reconciled.provenance ← BuildProvenanceChain(
        patient_records, conflicts)

    AuditLog.record(reconciled)
    RETURN reconciled
```

```
52: END FUNCTION
```

Complexity: O(M + P + A + D + K²) where M = medication entries, P = problem entries, A = allergy entries, D = documents, and K = total conflicts requiring negotiation. Independent agent processing (lines 9–16) executes in parallel.

B. Algorithm 2: Medication Reconciliation Agent

```
Input:  med_lists — Medication lists from N sources,
        each entry with code, name, dose, route,
        frequency, source_org, timestamp
Output: reconciled_meds — Unified medication list
1:  FUNCTION ReconcileMedications(med_lists)
2:      // Phase 1: Normalize all entries to RxNorm
3:      normalized ← []
4:      FOR EACH entry IN Flatten(med_lists) DO
5:          rxcui ← NULL
6:          // Cascading normalization
7:          IF entry.rxnorm_code EXISTS THEN
8:              rxcui ← entry.rxnorm_code
9:          ELSE IF entry.ndc_code EXISTS THEN
10:             rxcui ← NDCtoRxNorm(entry.ndc_code)
11:         ELSE
12:             rxcui ← FuzzyMatchRxNorm(
13:                 entry.drug_name, threshold=0.85)
14:         END IF
15:         entry.rxcui ← rxcui
16:         entry.norm_confidence ← GetNormConfidence(
17:             rxcui, entry)
18:         normalized.append(entry)
19:     END FOR
20:
21:     // Phase 2: Cluster by RxCUI
22:     clusters ← GroupBy(normalized, key=rxcui)
23:
24:     // Phase 3: Within-cluster deduplication
25:     reconciled_meds ← []
26:     FOR EACH (rxcui, entries) IN clusters DO
27:         IF Len(entries) == 1 THEN
28:             reconciled_meds.append(
29:                 entries[0].withConfidence(
30:                 entries[0].norm_confidence))
31:         ELSE
32:             // Compare dose/route/frequency
```

```
33:             groups ← ClusterByDoseRouteFreq(entries)
34:             FOR EACH group IN groups DO
35:                 // Select authoritative entry
36:                 best ← SelectByProvenance(group)
37:                 best.sources ← group.allSources()
38:                 best.agreement ← Len(group)
39:                     / Len(entries)
40:                 reconciled_meds.append(best)
41:             END FOR
42:         END IF
43:     END FOR
44:
45:     RETURN reconciled_meds
46: END FUNCTION
```

Complexity: O(M × log M) for normalization lookups, O(M) for clustering, O(C × E^2) for within-cluster deduplication where C = clusters, E = entries per cluster. Dominated by RxNorm API lookups, which are cached with LRU eviction.

C. Algorithm 3: Inter-Agent Negotiation Protocol

```
Input:  conflict — Detected cross-domain conflict
          agents — Involved reconciliation agents
Output: resolution — Resolved or escalated decision
1:  FUNCTION NegotiateResolution(conflict, agents)
2:      // Classify conflict type
3:      type ← ClassifyConflict(conflict)
4:
5:      // Gather evidence from each agent
6:      evidence_packages ← {}
7:      FOR EACH agent IN agents DO
8:          evidence_packages[agent] ← {
9:              decision: agent.getDecision(
10:                 conflict.entity),
11:             confidence: agent.getConfidence(
12:                 conflict.entity),
13:             terminology_evidence:
14:                 agent.getTerminologyLookups(
15:                 conflict.entity),
16:             provenance_evidence:
17:                 agent.getProvenanceData(
18:                 conflict.entity),
19:             domain_rules:
```

```
20:             agent.getApplicableRules(
21:             conflict.entity)
22:         }
23:     END FOR
24:
25:     // Apply resolution strategy by conflict type
26:     SWITCH type:
27:         CASE FACTUAL:
28:             winner ← ArgMaxConfidence(
29:                 evidence_packages)
30:             IF winner.confidence -
31:                 runner_up.confidence > 0.15 THEN
32:                 RETURN Resolved(winner.decision,
33:                     winner.confidence)
34:             ELSE
35:                 RETURN Escalate("Narrow confidence
36:                     margin")
37:             END IF
38:
39:         CASE SCOPE:
40:             primary ← GetPrimaryDomainAgent(
41:                 conflict.affected_domain)
42:             IF primary.confidence > 0.60 THEN
43:                 RETURN Resolved(primary.decision,
44:                     primary.confidence)
45:             ELSE
46:                 RETURN Escalate("Primary agent
47:                     low confidence")
48:             END IF
49:
50:         CASE PRIORITY:
51:             // Use DPA authority scoring
52:             authority ← DPA.getAuthorityScores(
53:                 conflict.sources)
54:             winner ← ArgMaxAuthority(
55:                 evidence_packages, authority)
56:             RETURN Resolved(winner.decision,
57:                 winner.confidence * authority.score)
58:
59:         DEFAULT:
60:             RETURN Escalate("Unknown conflict type")
61:     END SWITCH
62: END FUNCTION
```

Complexity: O(A × E) where A = agents involved (typically 2–3), E = evidence elements per agent. Completes in < 50 ms per conflict.

V. Results

A. Overall Reconciliation Accuracy

Table IV presents overall reconciliation accuracy across all 47,200 injected inconsistencies.

TABLE IV: OVERALL RECONCILIATION ACCURACY

Metric	B1: Rules	B2: Single-Model ML	B3: MARA (no negotiation)	MARA (full)	Target
Overall accuracy	71.8%	83.6%	89.7%	**92.4%**	> 90%
False merge rate	8.2%	4.1%	2.8%	**2.1%**	< 3%
False separation rate	14.7%	9.3%	7.2%	**5.8%**	< 10%
Human escalation rate	N/A	18.4%	16.2%	**11.3%**	< 15%
Processing time (per patient)	0.4 sec	3.1 sec	2.4 sec	**2.8 sec**	< 5 sec

MARA outperforms all baselines on every metric. Notably, the full MARA with negotiation improves accuracy by 2.7 percentage points over MARA without negotiation (B3), demonstrating the value of the inter-agent negotiation protocol. The rule-based baseline (B1) achieves the fastest processing but suffers from high error rates, particularly in false merges (8.2%) — clinically dangerous errors where distinct entities are incorrectly combined.

B. Per-Domain Accuracy

Table V reports accuracy by reconciliation domain.

TABLE V: PER-DOMAIN RECONCILIATION ACCURACY (MARA FULL)

Domain	Inconsist-encies	Correct	Accuracy	False Merge	False Sep
Patient Identity	4,400	4,224	96.0%	0.4%	3.6%
Medication Harmonization	17,200	16,357	95.1%	1.8%	3.1%
Problem List Deduplication	13,800	12,047	87.3%	3.4%	9.3%
Allergy Cross-Reference	5,400	4,968	92.0%	1.2%	6.8%
Document Provenance	6,400	5,984	93.5%	N/A	6.5%
All Domains	**47,200**	**43,580**	**92.4%**	**2.1%**	**5.8%**

Medication harmonization achieves the highest accuracy (95.1%) because RxNorm normalization provides a robust canonical representation: once two entries are mapped to the same RxCUI, deduplication is highly reliable. The fine-tuned BioBERT model handles free-text medication names with 91.3% normalization accuracy, capturing abbreviations, misspellings, and non-standard naming conventions.
Problem list deduplication has the lowest accuracy (87.3%) because SNOMED CT and ICD-10 express clinical conditions at different granularity levels, and the mapping between them is frequently one-to-many or imprecise. For example, ICD-10 code E11.65 ("Type 2 diabetes mellitus with hyperglycemia") and SNOMED CT concept 44054006 ("Diabetes mellitus type 2") are clinically related but not identical — whether to merge them depends on clinical context that is often unavailable from the coded data alone.
Allergy cross-reference achieves 92.0% accuracy but faces the unique challenge of allergy-versus-intolerance classification, where clinical sources genuinely disagree. ARA's conservative safety protocol (retaining the more severe classification) is clinically appropriate but counted as an accuracy miss when the ground truth labels the reaction as an intolerance.

C. Conflict Resolution Performance

Table VI details inter-agent conflict resolution.

TABLE VI: INTER-AGENT CONFLICT RESOLUTION PERFORMANCE

Conflict Type	Occurrences	Auto-Resolved	Resolution Rate	Accuracy of Auto-Resolution
Factual (duplicate disagreement)	1,847	1,512	81.9%	94.2%
Scope (cross-domain impact)	923	644	69.8%	91.7%
Priority (source authority)	1,134	912	80.4%	93.8%
All Conflicts	**3,904**	**3,068**	**78.6%**	**93.6%**

The negotiation protocol resolves 78.6% of inter-agent conflicts without human escalation, with 93.6% accuracy among auto-resolved conflicts. Scope conflicts have the lowest resolution rate (69.8%) because they involve genuinely ambiguous cross-domain interactions where automated resolution carries higher risk.

D. Agent-Pair Conflict Analysis

Table VII identifies which agent pairs generate the most conflicts, revealing the cross-domain interaction hotspots.

TABLE VII: CONFLICT FREQUENCY BY AGENT PAIR

Agent Pair	Conflict Count	Primary Conflict Type	Resolution Rate
MRA ↔ ARA	1,247	Scope (medication listed as allergen)	72.3%
MRA ↔ PRA	834	Factual (medication-implied condition)	83.1%
PRA ↔ DPA	712	Priority (document authority for problems)	81.6%
MRA ↔ DPA	623	Priority (medication list currency)	82.4%
ARA ↔ PRA	312	Scope (allergy-related conditions)	76.9%
PIA ↔ all agents	176	Factual (identity uncertainty propagation)	64.2%

The MRA↔ARA pair generates the most conflicts (1,247) because medications and allergies are clinically interlinked — a medication entry may appear in both the medication list and the allergy record, and the two agents may reach different conclusions about whether to merge or separate these entries. The lowest resolution rate (PIA↔all, 64.2%) reflects the cascading nature of identity uncertainty: when PIA is uncertain, all downstream agents inherit that uncertainty.

E. Normalization Performance

Table VIII reports the performance of the terminology normalization sub-systems that underpin agent accuracy.

TABLE VIII: TERMINOLOGY NORMALIZATION PERFORMANCE

Normalization Task	Method	Accuracy	Mean Latency
Medication → RxCUI (coded)	Direct lookup + NDC mapping	98.7%	2 ms
Medication → RxCUI (free-text)	Fine-tuned BioBERT	91.3%	45 ms
Problem → SNOMED CT (from ICD-10)	UMLS cross-reference	89.4%	8 ms
Problem → SNOMED CT (from free-text)	SNOMED CT search + semantic similarity	82.1%	62 ms
Allergy substance → RxNorm ingredient	RxNorm ingredient hierarchy	94.8%	5 ms
Document type classification	Rule-based + LOINC document codes	96.2%	1 ms

Free-text normalization is consistently the performance bottleneck, both in accuracy and latency. This reflects the fundamental challenge of unstructured clinical documentation: vendor-specific abbreviations, misspellings, and non-standard terminology create a long tail of entries that resist automated normalization.

F. Processing Performance

Table IX reports end-to-end processing performance.

TABLE IX: PROCESSING PERFORMANCE

Metric	Value
Mean processing time per patient (4 sources)	2.8 sec
Median processing time	2.1 sec
P95 processing time	6.4 sec
P99 processing time	11.2 sec
Throughput (patients/minute)	21.4
Agent parallelism utilization	89% (4 agents run concurrently)
RxNorm API cache hit rate	94.7%
UMLS cache hit rate	91.2%
Processing bottleneck	BioBERT inference (43% of total time)

The 2.8-second mean processing time is within the 5-second target, enabling real-time reconciliation during clinical data exchange. The P99 latency of 11.2 seconds occurs for patients with unusually large records (>20 medications, >15 problems from 4 sources), where BioBERT free-text normalization dominates processing time.

G. Provenance Impact Analysis

Table X demonstrates the impact of provenance-aware reconciliation by comparing MARA with and without provenance metadata from the integration layer [15].

TABLE X: IMPACT OF PROVENANCE METADATA ON RECONCILIATION ACCURACY

Configuration	Overall Accuracy	Authority Conflict Resolution	Correct Source Selection
MARA without provenance	89.1%	68.4%	74.2%
MARA with provenance	**92.4%**	**80.4%**	**88.7%**
Improvement	**+3.3 pp**	**+12.0 pp**	**+14.5 pp**

Provenance metadata contributes a 3.3 percentage point improvement in overall accuracy, with the largest impact on authority conflict resolution (+12.0 pp) and correct source selection (+14.5 pp). When agents must choose between conflicting information from different organizations, knowing the transformation history and source authority of each data element — provided by integration layer lineage tracking [15] — significantly improves decision quality.

VI. Discussion

A. Multi-Agent Advantage

The 8.8 percentage point accuracy improvement of MARA over the single-model baseline (B2: 83.6% → MARA: 92.4%) validates the multi-agent hypothesis: domain specialization enables each agent to embed deep knowledge (pharmacological, ontological, clinical) that a generalist model cannot effectively capture. The ablation between MARA with and without negotiation (B3: 89.7% → MARA: 92.4%) further demonstrates that inter-agent collaboration through the negotiation protocol adds significant value beyond independent specialization.

B. Clinical Safety Considerations

The 2.1% false merge rate is clinically significant: each false merge represents a case where two clinically distinct entities are incorrectly combined — potentially masking a genuine duplicate medication, losing a unique allergy record, or collapsing distinct clinical conditions. While MARA meets the 3% target, further reduction is essential for production deployment. The most dangerous false merges occur in medications (1.8%) where dose discrepancies between sources are incorrectly classified as documentation variations rather than genuine clinical differences.

The conservative allergy classification protocol — retaining the more severe classification — is a deliberate safety design choice that prioritizes avoiding missed allergies over perfect classification accuracy. In clinical practice, a false allergy label is an inconvenience (narrowed medication options) while a missed allergy is a potential life-threatening adverse event.

C. Scalability Considerations

MARA's 2.8-second per-patient processing time and 21.4 patients/minute throughput are sufficient for current exchange volumes. However, as TEFCA exchange volumes continue to grow [1], scalability will require: horizontal scaling of agent instances for parallel patient processing; pre-computed normalization caches with periodic refresh; and GPU-accelerated BioBERT inference pools for free-text normalization (currently the bottleneck at 43% of processing time).

D. Provenance as a Reconciliation Enabler

The 3.3 percentage point accuracy improvement from provenance metadata (Table X) demonstrates that data lineage tracking [15] is not merely an audit capability but an active reconciliation enabler. When field-level transformation provenance is available — documenting how each data element was derived from its source through integration engine processing — reconciliation agents can make more informed trust assessments. Specifically, provenance enables MARA to: (a) identify data elements that have undergone lossy vocabulary translation, reducing their authority weight; (b) distinguish original clinical documentation from derived or summarized entries; and (c) detect timestamp discrepancies caused by processing delays rather than clinical timing differences.

E. Limitations

1. **Synthetic data only:** The evaluation uses Synthea-generated records with programmatically injected inconsistencies. Real-world cross-organizational data exhibits messier, more ambiguous patterns that may reduce accuracy. Production validation is essential.
2. **English-language bias:** All terminology normalization and NLP components are trained on English-language clinical text. Multi-lingual healthcare environments require language-specific agent training.
3. **Four-organization limit:** Evaluation covers four contributing organizations. Reconciliation complexity grows combinatorially with source count; accuracy with 8+ sources requires separate validation.
4. **BioBERT fine-tuning data:** The medication and problem normalization models are fine-tuned on publicly available clinical NLP corpora. Organization-specific vocabularies and abbreviations may require local fine-tuning.
5. **Static agent configuration:** Agent confidence thresholds and negotiation parameters are manually configured. Adaptive threshold learning from reconciliation outcomes is noted for future work.
6. **No longitudinal reconciliation:** MARA reconciles point-in-time snapshots. Tracking how a patient's reconciled record evolves over time (longitudinal reconciliation) introduces additional complexity not addressed.

F. Threats to Validity

Construct validity: Programmatic inconsistency injection follows observed patterns from production HIEs but cannot guarantee exhaustive coverage of real-world variability. Ground truth labels are generated programmatically, introducing potential labeling bias.

Internal validity: Agent training data and evaluation data are generated from the same Synthea pipeline, creating potential distributional similarity. Cross-corpus evaluation would strengthen findings.

External validity: Single synthetic patient population. Cross-demographic and cross-geographic validation needed.

Reliability: Published Synthea seeds, terminology service versions, and model checkpoints enable full reproducibility.

VII. Conclusion and Future Work

This paper presented MARA — a Multi-Agent Reconciliation Architecture for cross-organizational clinical data reconciliation. Five domain-specialized AI agents — for patient identity, medications, problems, allergies, and document provenance — collaborate through a structured negotiation protocol to produce unified, confidence-scored patient records from disparate organizational sources. Evaluation demonstrates 92.4% overall reconciliation accuracy with 95.1% in medication harmonization and 87.3% in problem list deduplication. The negotiation protocol resolves 78% of inter-agent conflicts without human escalation, and end-to-end processing averages 2.8 seconds per patient.

MARA demonstrates that multi-agent specialization with collaborative negotiation significantly outperforms both rule-based and single-model approaches for cross-organizational clinical data reconciliation. Provenance-aware reconciliation leveraging integration layer lineage metadata [15] provides a 3.3 percentage point accuracy improvement, establishing data lineage as an active enabler of reconciliation quality.

Future work includes: (1) production validation with de-identified cross-organizational clinical data; (2) reinforcement learning for adaptive confidence threshold optimization based on reconciliation outcomes; (3) additional agents for laboratory result reconciliation and vital sign harmonization; (4) longitudinal reconciliation tracking how patient records evolve over time across organizations; (5) federated agent training across organizations without centralizing protected health information; (6) integration with clinical decision support systems to surface reconciliation uncertainties at the point of care; and (7) expansion to support non-English clinical documentation.

REFERENCES:

[1] ONC, "Trusted Exchange Framework and Common Agreement (TEFCA)," U.S. DHHS, 2024.
[2] Carequality, "Carequality Connected Agreement," 2023. [Online]. Available: https://carequality.org/
[3] D. Bender and K. Sartipi, "HL7 FHIR: An agile and RESTful approach to healthcare information exchange," in *Proc. IEEE CBMS*, 2013, pp. 326–331.
[4] S. T. Rosenbloom et al., "Data from clinical notes: A perspective on the tension between structure and flexible documentation," *JAMIA*, vol. 18, no. 2, pp. 181–186, 2011.
[5] R. Haux, "Health information systems — past, present, future," *Int. J. Med. Inform.*, vol. 75, pp. 268–281, 2006.
[6] The Joint Commission, "National Patient Safety Goals: Medication Reconciliation," 2023.
[7] C. Dolin et al., "HL7 Clinical Document Architecture, Release 2," *JAMIA*, vol. 13, no. 1, pp. 30–39, 2006.
[8] E. E. Gleason et al., "Results of the Medications at Transitions and Clinical Handoffs (MATCH) study: An analysis of medication reconciliation errors and risk factors at hospital admission," *J. General Internal Medicine*, vol. 25, no. 5, pp. 441–447, 2010.
[9] M. Wooldridge, *An Introduction to MultiAgent Systems*, 2nd ed. John Wiley & Sons, 2009.
[10] A. Isern and A. Moreno, "A systematic literature review of agents applied in healthcare," *J. Medical Systems*, vol. 40, no. 2, pp. 1–14, 2016.
[11] R. K. Srihari, "Intelligent multi-agent systems for healthcare: Architecture and applications," in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, 2005, pp. 3012–3017.
[12] G. Cabac et al., "Modeling multi-agent-based clinical workflows," in *Proc. Int. Workshop on Agent-Applied Systems*, 2006, pp. 44–58.
[13] HL7 International, "HL7 v2.x Messaging Standard," 2019.
[14] M. Eichelberg et al., "A survey and analysis of electronic healthcare record standards," *ACM Computing Surveys*, vol. 37, no. 4, pp. 277–315, 2005.
[15] S. Sundaram, "Managing clinical data lineage in distributed healthcare integration environments: A metadata instrumentation framework for end-to-end provenance tracking," *IJAIDSML*, vol. 7, no. 1, pp. 373–380, Mar. 2026, doi: 10.63282/3050-9262.IJAIDSML-V7I1P159.
[16] S. Nelson et al., "Normalized names for clinical drugs: RxNorm at 6 years," *JAMIA*, vol. 18, no. 4, pp. 441–448, 2011.
[17] O. Bodenreider, "The Unified Medical Language System (UMLS): Integrating biomedical terminology," *Nucleic Acids Research*, vol. 32, no. suppl_1, pp. D267–D270, 2004.

[18] A. R. Hevner et al., "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.
[19] J. Walonoski et al., "Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record," *JAMIA*, vol. 25, no. 3, pp. 230–238, 2018.
[20] J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020