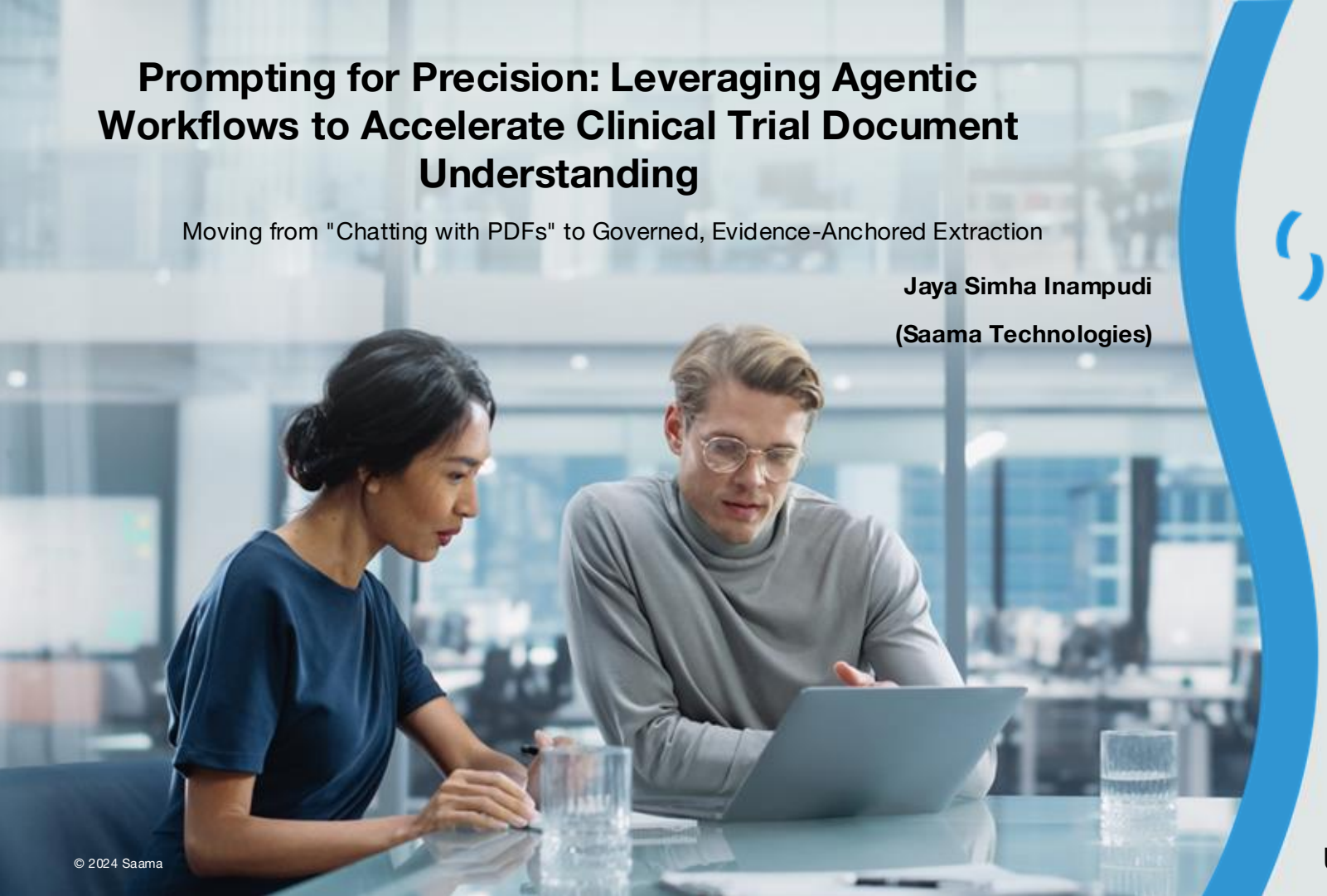
Prompting for Precision: Leveraging Agentic Workflows to Accelerate Clinical Trial Document Understanding

Moving from "Chatting with PDFs" to Governed, Evidence-Anchored Extraction

Jaya Simha Inampudi

(Saama Technologies)



US Connect 2026

The Challenge: The "Needle in the Haystack"

The Daily Reality for Clinical Programmers

High-Stakes Documentation: Success hinges on mastering complex documents like Protocols, SAPs, and Standards.

Operational Friction: It is not just "reading"; it is repeatedly answering precise operational questions:

- *What is the precise TEAE window?*
- *Where are the censoring rules for the primary endpoint?*
- *How does alpha recycling work in the multiplicity hierarchy?*

The Cost: Answers are currently driven by human memory, leading to inconsistency, slow onboarding, and significant time drain.

Why Naïve AI Fails in Pharma

"Chatting with Documents" is Dangerous in Regulated Work

- **Hallucinations**: Generative models fill missing details with "plausible" conventions (e.g., inventing a 30-day safety window).
- **Context Drift**: Models conflate similar terms, confusing "Primary Endpoint" definitions with "Primary Estimand" derivations.
- **Lost Provenance**: Answers cannot be tied back to a specific line in the source text, making QC impossible.
- **Non-Determinism**: Run the same prompt twice, get different answers.

The Verdict: A response is only useful if it is faithful, consistent, and traceable .

The Core Thesis

Shifting the Paradigm To use LLMs safely, we must move from "Prompting as a Query" to "Prompting as a Governed Extraction Workflow."

The Strategy:

1. **Scope Discipline**: Agents only extract what they "own" (no cross-domain leakage).
2. **Structured Outputs**: Stable JSON/YAML schemas, not free text.
3. **Evidence Linkage**: Every object must reference a specific source chunk.
4. **Gap Surfacing**: Missing info is flagged as a "Gap," not filled with a guess.

The Solution Architecture

Separating Structure from Semantics

Layer 1: Deterministic Structural Layer (No LLM)

- **Indexer**: Segments documents using structure cues (TOC, headers), ignoring unstable pagination.
- **Role Labeler**: Routes chunks to "Schema Owners" (e.g., Stats, Safety, Design).

Layer 2: LLM Semantic Extraction Layer

- **Scoped Agents**: Specialized clusters (e.g., Cluster C for Stats) extract only their specific subtree.

Layer 3: Governance & Assembly

- **Orchestrator**: Merges outputs and builds a "UID Graph" to check references.
- **Auditor**: Cross-checks traceability and flags coverage gaps.

Deep Dive – Deterministic Indexing

Solving the "Needle in the Haystack"

- **Problem:** SAPs are messy. They have inconsistent headers, OCR artifacts, and repeated terms (e.g., "Safety" appears in Objectives, Methods, and Reporting).
- **Solution:** Anchor-Based Segmentation. We do not rely on page numbers. We use stable structural anchors (Section IDs + Titles) to create a robust map.
- **Benefit:** This bounds the LLM's context. When asking about "Statistical Models," the LLM only sees the Statistical Analysis section, preventing contamination from the Introduction.

Deep Dive – The "Ownership" Model

Solving Context Drift

- **Role Labeling:** Instead of generic topic tagging, we assign ownership based on the Output Schema.
  - *Cluster D (Strategy) owns the Multiplicity Hierarchy.*
  - *Cluster E (Reporting) references the Hierarchy but cannot redefine it.*

- **Result:** A single canonical source for every definition. "Semantic Drift" is eliminated because downstream agents are forced to use the upstream definition.

Deep Dive – Constrained Extraction

Solving Hallucinations via "Prompt Contracts"

- **The Rules:**
  1. **Strict Schema:** Output must fit a rigid JSON shape.
  2. **Null Discipline:** If the text is not explicit, output null. Do not infer.
  3. **Evidence Requirement:** Every value must have a sap_ref (pointer to source).

- **The Outcome:** The LLM behaves like a Schema-Constrained Parser, not a creative writer.

Governance Layer – The UID Graph

Ensuring Clinical Integrity

Clinical data is relational:

- *Endpoint -> Estimand -> Analysis -> TFL Output.*

The UID Graph: The Orchestrator builds a graph of all extracted entities to ensure referential integrity.

- *Check:* Does the Analysis reference an undefined Population?
- *Check:* Is an Endpoint listed in the TFLs but missing from the Estimands?

Micro-Loops: If a definition is missing, the system triggers a targeted, bounded re-read to find it—avoiding infinite loops .

Proof of Value – Before & After

Ensuring Clinical Integrity

TEAE Definition Query (Safety)

Before (naïve chat-with-document):

User: What is the TEAE definition in this SAP?

Model answer (typical failure pattern): TEAEs are adverse events occurring from first dose until 30 days after last dose, including events that worsen from baseline.

Issues:

- The model selected “30 days” without evidence.
- No source reference.
- “Worsen from baseline” may be defined differently in the SAP.

After (precision-governed extraction style):

```
safety.teae_definition:
  description: "Treatment emergent adverse events are defined relative to the date of first dose [Week 0 (Visit 3) unless indicated otherwise]."
  lag_time: null
  start_rules: "events with a start date that is equal to or greater than the date of first dose; events that start prior to the date of first dose and worsen after that date; events that start and resolve prior to the date of first dose, but then recur after that the date of first dose."
  stop_rules: null
  worsening_baseline_logic: "events that start prior to the date of first dose and worsen after that date"
  sap_ref: "Sec 11.2"
flags:
  - type: "GAP_SAFETY"
    code: "MISSING_TEAE_LAG_TIME"
    schema_path: "safety.teae_definition.lag_time"
    sap_ref: "Sec X.Y (Safety Definitions)"
    notes: "TEAE lag time after last dose is not explicitly stated in the routed safety text.“
```

What changed:

- The system refuses to invent “30 days.”
- It returns a structured partial extraction and creates a clear gap for follow-up.

Proof of Value – Before & After

Ensuring Clinical Integrity

Endpoint ↔ Analysis ↔ TLF Alignment

Before (naïve summarization):

"Primary endpoint OS is analyzed using Cox model and displayed in Table 14.1."

Issues:

- "Table 14.1" may not be correct, and OS may be displayed in a figure, not a table.
- Missing population context.
- Missing evidence references.
- No link structure to validate completeness

After (governed, linked output view):

```
endpoints.items[0]:
  endpoint_uid: "END-EXAMPLE-01"
  name: "Overall Survival"
  class: "Primary"
  type: "TTE"
  definition: "Time from randomization to death from any
cause."
  analysis_population_uids: ["POP-EXAMPLE-01"]
  sap_ref: "Sec A.B (Endpoints)"

analyses.items[0]:
  analysis_uid: "ANA-EXAMPLE-01"
  name: "Primary OS model"
  endpoint_uid: "END-EXAMPLE-01"
  population_uid: "POP-EXAMPLE-01"
  model:
    method: "Cox proportional hazards"
    formula: null
  sap_ref: "Sec C.D (Primary Analysis)"
```

```
tlf_specifications.items[0]:
  tlf_uid: "OUT-EXAMPLE-01"
  output_id: "Figure F.X"
  type: "Figure"
  title: "Kaplan–Meier Plot of Overall Survival"
  population_uid: "POP-EXAMPLE-01"
  endpoint_uids: ["END-EXAMPLE-01"]
  analysis_uids: ["ANA-EXAMPLE-01"]
  sap_ref: "Annex Z (Shells)"

uid_references:
  - uid: "END-EXAMPLE-01"
    category: "endpoint"
    used_in_path: "analyses.items[0].endpoint_uid"
    sap_ref: "Sec C.D"
  - uid: "ANA-EXAMPLE-01"
    category: "analysis"
    used_in_path:
"tlf_specifications.items[0].analysis_uids[0]"
    sap_ref: "Annex Z"
```

What changed:

- The system produces a consistent link graph across endpoint, analysis, and output.
- Each object is anchored to a location (sap_ref).
- QC can now automatically detect if an output references an undefined analysis or endpoint.

Summary

Prompting for Precision

- LLMs can accelerate clinical workflows, but only with **Governance**.
- We must shift from "Summarization" to **"Traceable, Structured Interpretation."**

Key Takeaway: By combining deterministic indexing with agentic constraints, we achieve the **speed of AI** with the **rigor of Clinical Programming**.

Jaya Simha Inampudi
Sr Director,
Biometrics & Clinical Solutions (BCS), Saama
jaya.inampudi@saama.com