Paper ML06

The AI-Powered Scribe: An Agentic Framework for Automated CSR Authoring

Author: Allwyn Dsouza (Saama Technologies)

ABSTRACT

A multi-agent AI system designed to automate and streamline the creation of Clinical Study Report (CSR). Our framework deploys a team of specialized AI agents, each tasked with authoring a specific section of the CSR by intelligently sourcing information from relevant documents like the protocol, SAP, and TLFs.

The agents collaborate to ensure a consistent and coherent narrative throughout the report. Central to this process is a robust Human-in-the-Loop (HITL) governance model, allowing clinical experts to review, refine, and approve all AI-generated content. This ensures complete scientific accuracy and regulatory compliance.

This agent-driven approach transforms the complex authoring process, enhancing efficiency and consistency.

INTRODUCTION

Clinical Study Reports represent the culmination of years of clinical research, synthesizing protocol intent, trial conduct, statistical analysis, and clinical interpretation into a single authoritative regulatory document. Despite advances in standards such as CDISC SDTM and ADaM, and despite widespread adoption of automated table and listing generation, CSR development continues to be characterized by long timelines, repeated review cycles, and significant late-stage rework.

One reason for this persistent inefficiency is that CSR development is often treated as a downstream documentation activity rather than an integrated scientific process. Medical writers typically begin drafting once statistical outputs are finalized, only to discover during review that interpretations differ across sections, assumptions were misunderstood, or traceability to underlying analyses is incomplete. These issues are then resolved through iterative review cycles involving statisticians, clinicians, quality reviewers, and regulatory stakeholders.

These challenges are often amplified by scale. Teams may be responsible for dozens of concurrent studies across multiple therapeutic areas, each with unique protocols and timelines. Under such conditions, incremental improvements in document templates or isolated automation tools provide limited relief. A more fundamental rethinking of CSR generation as a coordinated, system-level activity is required.

JOURNEY TOWARDS AGENTIC AI

The rapid adoption of generative AI has led many organizations to experiment with applying large language models directly to CSR drafting. In its simplest form, this approach involves prompting a model with statistical outputs or summary text and asking it to generate narrative sections. While such experiments often produce fluent prose, they quickly encounter limitations when applied to real-world CSR workflows.

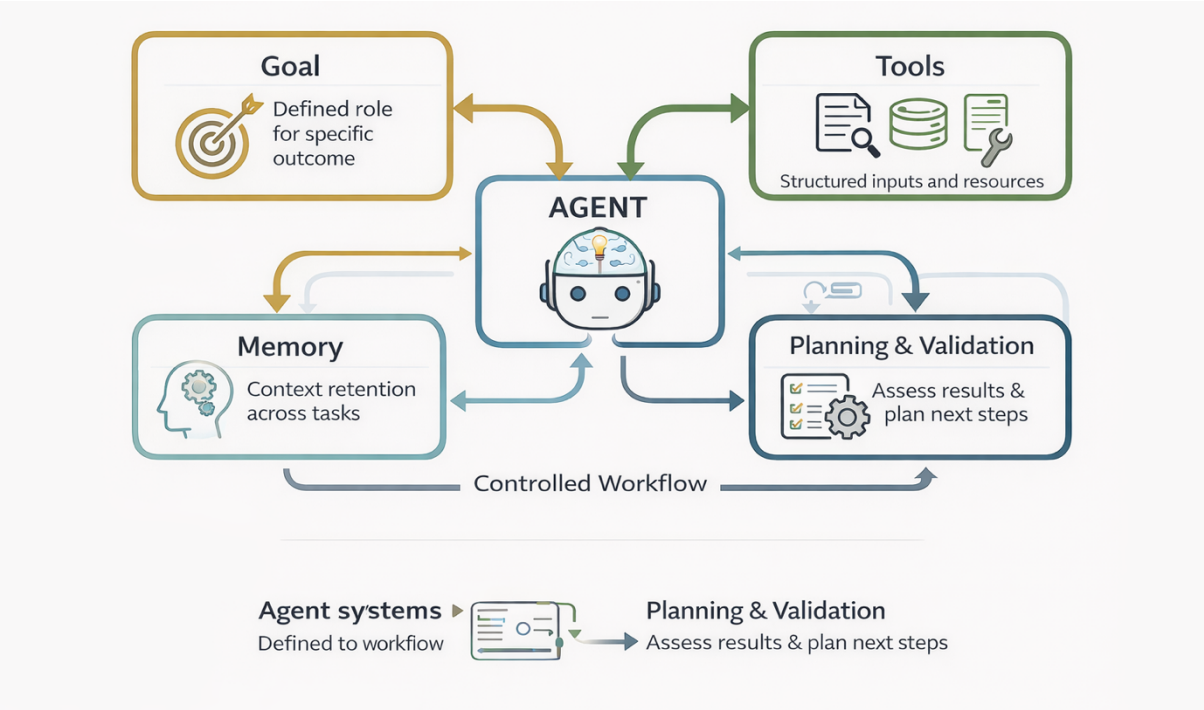
Prompt-based systems struggle to maintain consistency across sections, particularly when the same endpoint or population is described in multiple contexts such as efficacy results, safety summaries, and benefit–risk discussions. They also lack persistent memory, making it difficult to ensure that interpretations agreed upon in one section are respected elsewhere in the document. Most importantly, these systems

provide little insight into how conclusions were derived, making validation and regulatory acceptance challenging.

Agentic AI represents a shift away from this paradigm. Instead of relying on a single model to generate large portions of text, agentic systems decompose the CSR generation task into a network of interacting agents, each with a narrowly defined responsibility. This architectural shift reflects the reality that CSR development is not a single cognitive task, but a coordinated set of activities involving interpretation, verification, and decision-making.

WHAT CONSTITUTES AN AGENT IN PRACTICE

In practical implementations, an agent is more than a wrapper around a language model. It is a structured component designed to perform a specific role reliably within a controlled workflow.



Each agent operates with a defined **goal,** such as interpreting the primary efficacy endpoint or verifying consistency between narrative text and a corresponding table. This goal constrains the agent's behaviour and limits its scope, reducing the risk of unintended outputs.

Agents are also equipped with **tools** appropriate to their role. For example, a statistical interpretation agent may have access to structured metadata describing the statistical model, population definitions, and hypothesis testing framework, rather than raw tables alone. By working with structured inputs, the agent can reason about results more consistently and transparently.

Memory is another critical feature. Unlike stateless prompt-based systems, agents maintain contextual memory that allows them to retain relevant facts across interactions. For instance, once an agent has determined that the primary endpoint was met in the intent-to-treat population, this conclusion can be stored and reused consistently across sections.

Finally, agents incorporate **basic planning and validation logic**, enabling them to assess whether their outputs meet predefined acceptance criteria

MULTI-AGENT ARCHITECTURE FOR CSR GENERATION

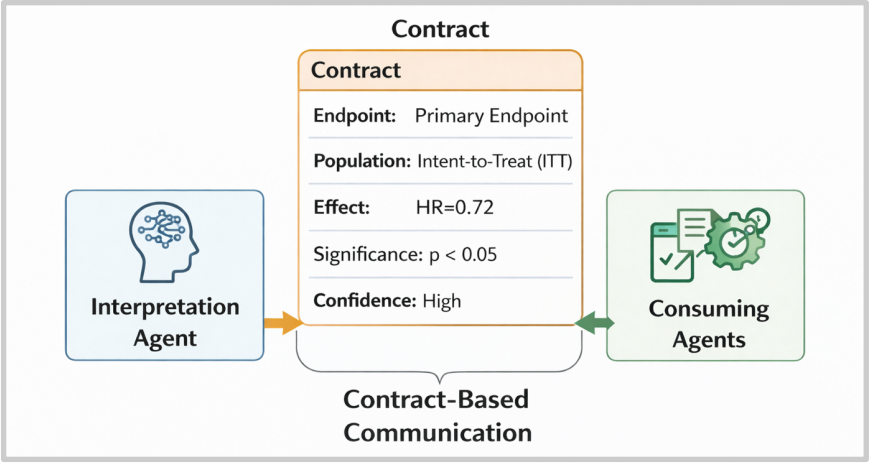
CSR development mirrors a multi-disciplinary human workflow. Statisticians interpret data, medical writers translate findings into narrative form, quality reviewers check for consistency, and regulatory experts ensure compliance. Attempting to replicate all of these roles within a single AI agent leads to systems that are difficult to reason about, test, or govern.

A multi-agent architecture addresses this issue by assigning discrete responsibilities to specialized agents. For example, a Study Context Agent focuses exclusively on ingesting protocol and SAP content, including amendments and deviations. Its role is to establish and maintain a consistent understanding of study intent that downstream agents can reference. A Statistical Interpretation Agent then uses this context to interpret analysis outputs, producing structured summaries of findings without generating narrative prose.

Narrative generation is handled by a separate agent, which consumes these structured interpretations and applies regulatory writing conventions to produce text. Additional agents perform consistency checks across sections, manage traceability links, or oversee orchestration. This separation of concerns improves robustness and allows individual agents to be validated independently.

AGENT COMMUNICATION AND CONTRACT-BASED INTERACTIONS

In human teams, miscommunication is a major source of rework. The same principle applies to agentic systems.



Allowing agents to communicate via free-form natural language introduces ambiguity and complicates validation. To address this, agentic CSR systems rely on structured message passing, often referred to as contract-based communication.

For example, when a statistical interpretation agent completes its analysis of a primary endpoint, it may produce a structured message specifying the endpoint name, population, direction and magnitude of effect, statistical significance, and confidence level. This message serves as a contract that downstream agents can consume without reinterpretation. If assumptions change or discrepancies are detected, the contract can be updated explicitly, triggering downstream updates.

This approach not only improves reliability but also supports auditability. Each contract can be logged, reviewed, and traced back to its source, providing transparency into how conclusions were formed.

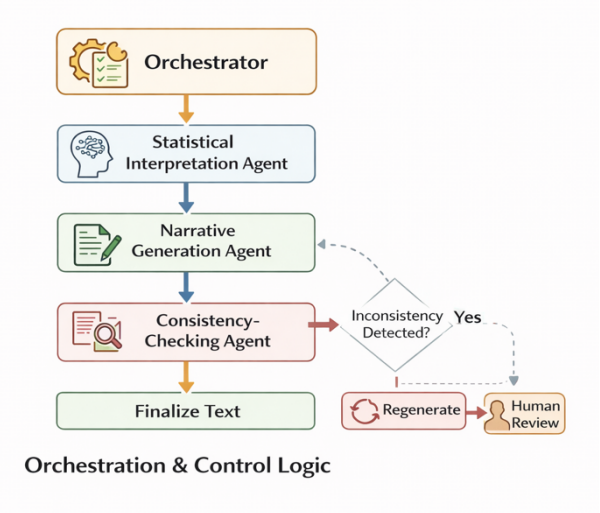
ORCHESTRATION AND CONTROL LOGIC

At the heart of a multi-agent system lies the orchestrator, which manages execution flow and ensures that agents operate within defined boundaries. The orchestrator does not generate scientific content; instead, it applies deterministic logic to decide which agents should run, in what order, and under what conditions.

For instance, during generation of an efficacy section, the orchestrator may first invoke the statistical interpretation agent, then pass its output to the narrative generation agent.

Once text is generated, a consistency-checking agent may be invoked to compare the narrative against the source tables. If inconsistencies are detected, the orchestrator can trigger a regeneration step or escalate the issue to a human reviewer.

This explicit control logic is essential in GxP environments, where opaque or autonomous behaviour is unacceptable. By making orchestration rules transparent and configurable, organizations can align system behaviour with regulatory expectations.



Orchestration & Control Logic

MEMORY AND CONTEXT MANAGEMENT

Memory plays a central role in ensuring coherence and reducing hallucination in agentic systems. In CSR generation, memory is often organized into multiple layers. Short-term memory captures the immediate context of the task at hand, such as the specific endpoint or population being discussed. Long-term memory stores study-wide facts and interpretations that must remain consistent across sections. Immutable memory records approved decisions that should not be altered without explicit authorization.

For example, once a benefit–risk conclusion has been approved by a clinical lead, it may be stored in immutable memory. Downstream agents can reference this conclusion but cannot modify it, ensuring consistency and preserving accountability. This layered approach to memory management supports both flexibility and control.

TRACEABILITY AS A GRAPH-BASED SYSTEM

Traditional CSR workflows treat traceability as a documentation requirement addressed late in the process. In contrast, agentic systems embed traceability directly into the generation workflow. Each narrative paragraph can be linked to the specific tables, datasets, and SAP sections that support it, forming a graph of evidence relationships.

This graph-based representation enables powerful capabilities. For example, if an analysis is updated, the system can identify all affected narrative sections automatically. Reviewers can navigate from a paragraph directly to its supporting evidence, improving transparency and review efficiency. In this way, traceability becomes an active system feature rather than a static checklist item.

OPPORTUNITIES ENABLED BY AGENTIC SYSTEMS

When CSR generation is implemented as a multi-agent system rather than a linear document workflow, a range of new operational possibilities emerge that are difficult or impossible to achieve using traditional approaches.

One of the most significant opportunities is the ability to **decouple narrative development** from final analysis lock. For example, an efficacy narrative for the primary endpoint can be drafted as soon as the statistical interpretation agent determines that the endpoint has been met in the intent-to-treat population, even if sensitivity analyses are still ongoing. When final analyses are available, the system can automatically update affected paragraphs and re-run consistency and traceability checks.

Another important opportunity lies in **early inconsistency detection**. In traditional workflows, discrepancies between tables, figures, and narrative text are often discovered during formal QC or regulatory review, when changes are costly and time-consuming. In an agentic system, consistency-checking agents can continuously monitor alignment between statistical outputs and narrative drafts. For instance, if a hazard ratio reported in a table differs from the value described in the text, the system can flag the discrepancy immediately and either trigger regeneration or escalate to a human reviewer.

Agentic systems also enable **role-specific and audience-specific views** of CSR content. Regulators, clinicians, statisticians, and quality reviewers often seek different levels of detail and explanation. By maintaining structured interpretations and traceability links, the system can support interactive exploration of CSR content. A reviewer could, for example, select a paragraph describing a safety finding and request an explanation of how the conclusion was derived, including links to supporting tables, datasets, and SAP sections.

NEW CHALLENGES AND ENGINEERING TRADE-OFFS

While agentic systems offer significant advantages, they also introduce new challenges that must be addressed deliberately.

- **Agent drift,** where an agent's behaviour changes over time due to evolving prompts, memory updates, or model versions. In a CSR context, even small changes in interpretation logic can have significant downstream effects, potentially altering narrative conclusions or traceability relationships.

  To mitigate this risk, agentic systems require explicit versioning and lifecycle management. Each agent, including its underlying model configuration, rules, and prompt templates, should be version-controlled in the same manner as traditional software components. This allows organizations to reproduce prior outputs and understand the impact of system changes.

- **Memory management** introduces additional trade-offs. While persistent memory is essential for maintaining consistency, uncontrolled memory growth or corruption can lead to unintended behaviour. For example, if an agent incorrectly stores an unapproved interpretation as a study-wide fact, that error may propagate across multiple sections. To address this, memory access must be carefully governed, with clear rules about which agents can write to long-term or immutable memory and under what conditions.

- **Over-automation** - As agentic systems become more capable, there may be a temptation to defer human involvement until late in the process. In regulated environments, this is undesirable. Human judgment must remain central, particularly for interpretation, benefit–risk assessment, and regulatory positioning. Successful implementations therefore treat automation as a means of surfacing decisions earlier, not replacing them.
- **System complexity** itself becomes a consideration. Multi-agent architectures are inherently more complex than single-agent or template-based systems. This complexity must be justified by clear benefits in scalability, quality, or governance. Organizations should therefore adopt agentic approaches incrementally, focusing first on high-impact use cases such as efficacy narratives or traceability management.

GXP GOVERNANCE

From a GxP perspective, the introduction of agentic AI does not eliminate the need for governance; rather, it reshapes it. Regulators are less concerned with whether AI is used and more concerned with how decisions are made, documented, and controlled. In an agentic CSR system, governance must therefore focus on transparency, accountability, and reproducibility.

Human-in-the-loop controls are a central requirement. While agents may generate interpretations or draft narratives, final approval must rest with qualified human reviewers. These approval points should be explicit and enforced by the system, ensuring that no content progresses to a finalized state without appropriate oversight.

Every agent action—whether generating an interpretation, updating a narrative, or flagging an inconsistency—should be logged with sufficient metadata to reconstruct the decision path. This includes the agent version, input data references, confidence assessments, and downstream effects. Such audit trails support both internal quality assurance and external regulatory inspection.

When implemented with these controls, agentic systems can meet GxP expectations while offering greater transparency than many traditional manual workflows.

IMPLICATIONS FOR MEDICAL WRITING

The adoption of agentic AI has important implications for how work is organized across medical writing, statistics. Rather than replacing roles, agentic systems reallocate effort toward higher-value activities. Medical writers spend less time reconciling numbers and more time focusing on scientific interpretation and narrative coherence. Statisticians shift from responding to ad hoc clarification requests to defining analytical intent and validation rules that agents can apply consistently.

Instead of scaling delivery by increasing headcount, organizations can scale by improving system capability. At the same time, new skills become important. Teams must develop capabilities in agent configuration, orchestration design, and governance. This evolution mirrors earlier transitions, such as the adoption of statistical programming standards or automated reporting frameworks, and represents a natural progression in the maturation of clinical development operations.

CONCLUSION

Clinical Study Report generation has long been treated as a documentation problem, addressed through templates, manual coordination, and late-stage review. This paper has argued that CSR development is more accurately understood as a scientific coordination problem, involving interpretation, validation, and alignment across multiple domains. Prompt-based generative AI addresses only a narrow slice of this challenge.

Agentic AI offers a fundamentally different approach. By decomposing CSR generation into a system of specialized, collaborating agents operating under explicit orchestration and governance, organizations can achieve scalable automation while preserving human accountability and scientific rigor. Importantly, this approach does not depend on fully autonomous behaviour; rather, it emphasizes transparency, traceability, and human-in-the-loop decision-making.

While this paper has focused on CSRs, the same agentic principles can be applied across the clinical development lifecycle. Similar architectures can support protocol authoring, SAP development, Integrated Summary of Safety (ISS) and Integrated Summary of Efficacy (ISE) generation, regulatory briefing documents, and even post-marketing safety reporting. Beyond regulatory documents, agentic systems can be applied to real-world evidence synthesis, clinical trial disclosure, and inspection readiness activities.

In each of these domains, the key insight remains the same: meaningful automation arises not from generating more text, but from engineering systems that support scientific reasoning, coordination, and accountability at scale. Agentic AI provides a promising foundation for this next phase of transformation in clinical development.

CONTACT INFORMATION

Author - Allwyn Dsouza

Email - allwyn.dsouza@saama.com

Company - Saama Technologies

Website - www.saama.com