POSITION BRIEF

Clinical Trial Protocol Automation

An Agentic Knowledge Orchestration Approach

Satya Suman Saridae

Director of Generative AI & Engineering, GHX

Technical Co-founder, F0rty2.ai

April 2026

This document outlines a perspective on clinical trial protocol automation shaped by nearly two decades across life sciences, health-tech, and enterprise AI — including prior work at Novartis building regulated AI systems for pharmacovigilance and enterprise GenAI infrastructure.

1. Relevant Experience at Novartis

1.1 DSUR Automation: Where the Problem Became Clear

The earliest and most formative experience was working on DSUR (Development Safety Update Report) automation. A DSUR, per ICH E2F, is an annual regulatory document synthesizing the safety profile of an investigational product across all ongoing clinical trials. It aggregates data from multiple sources — safety databases, line listings, literature surveillance, and cumulative subject exposure — into a structured narrative for regulatory submission.

The core challenge was multi-source summarization: pulling structured data from safety databases, unstructured narratives from case reports, and tabular data from exposure calculations, then producing a coherent document that met both scientific rigor and regulatory template requirements. This was not a pure AI problem or a pure engineering problem — it required both. The summarization pipeline had to handle heterogeneous data formats, maintain source traceability, and produce output that pharmacovigilance physicians could review and sign off on.

This experience established a pattern that has informed every subsequent system: structured document generation where AI aggregates evidence from disparate sources, applies a regulatory template, generates traceable narrative, and passes to a human expert for judgment.

1.2 AE Brain: Validating AI Under GxP

AE Brain was built on BioBERT and BERT embeddings for adverse event intelligence — processing safety data from clinical trials and supporting pharmacovigilance signal detection. This was in an era before retrieval-augmented generation was a standard pattern; the system relied on fine-tuned transformer embeddings for semantic matching and classification.

The technical challenge was significant, but the harder challenge was validation. Getting an AI system built on neural embeddings through GxP compliance under ICH-GCP and 21 CFR Part 11 required a full Computer System Validation (CSV) lifecycle: Installation Qualification, Operational Qualification, Performance Qualification, change control procedures, audit trail implementation, and electronic signature workflows. This also meant collaborating closely with the BSO (Business System Owner) and the Security Architecture team during external audit cycles — ensuring the system met not just functional validation requirements but also data security, access control, and audit readiness standards expected by external auditors.

This experience is directly transferable to protocol automation: any AI-assisted protocol authoring tool at a pharma company will face the same validation and security requirements. Understanding how to scope a system as GxP-relevant (advisory, human-in-the-loop) versus GxP-critical, how to structure the validation approach, and how to prepare for external audit scrutiny is practical knowledge that significantly reduces time-to-deployment.

1.3 GAIN: Enterprise GenAI Platform (CAS Team)

GAIN (Generative AI @ Novartis) was built by the CAS (Cognitive Automation Services) team in 2024, when retrieval-augmented generation was still an emerging pattern. The platform was designed as enterprise-grade GenAI infrastructure serving multiple functions across Novartis — from literature synthesis for drug discovery teams to document intelligence for regulatory affairs.

The architecture integrated document ingestion (PDFs, clinical study reports, regulatory filings, SOPs), embedding-based retrieval with source citation, and role-based access controls aligned with Novartis's existing data architecture. GAIN was production infrastructure: enterprise SSO, audit logging, and integration with the broader Novartis technology ecosystem.

The progression from DSUR automation to AE Brain to GAIN reflects a natural arc: solving a specific document generation problem, then building a validated AI system for a regulated use case, then scaling the underlying patterns into enterprise infrastructure. Each stage informed the next.

1.4 Current Work and Relevance to This Role

Since Novartis, my focus has been on building healthcare AI systems end-to-end: at GHX as Director of Generative AI & Engineering leading a team building therapeutic AI solutions for hospital networks, and independently through F0rty2.ai, developing a multi-agent Patient Relationship Manager for clinical settings. These roles required end-to-end ownership — from data infrastructure and model architecture through deployment and clinical workflow integration.

Protocol automation sits at the intersection of these experiences: regulatory document generation (DSUR), validated AI in pharma (AE Brain), enterprise RAG architecture (GAIN), and multi-agent workflow orchestration (F0rty2.ai). The problem space is well-suited to this combined background.

2. The Problem: Protocol Authoring as the Root Cause

Clinical trial protocols are the root cause document in pharmaceutical R&D. A well-designed protocol minimizes downstream costs; a poorly designed one cascades into enrolment delays ($2.25M in under-enrolling site expenses), protocol amendments ($450K each), and in the worst case, trial failure. Industry data shows 86% of trials do not meet enrolment timelines, and protocol design complexity has been trending upward across all design variables.

The core challenges that Novartis and the broader industry have identified in this space include:

Challenge	Description	Automation Opportunity
Semantic retrieval of historical trials	When designing a protocol, authors need to find the most relevant historical trials based on unstructured text (study title, outcomes, criteria). Current approaches rely on manual search.	Embedding-based semantic search with hybrid retrieval (BM25 + dense) and domain-specific reranking.
Enrolment duration prediction	Overly restrictive I/E criteria lead to enrolment delays. Authors lack data-driven feedback on how criteria choices impact enrolment timelines.	Predictive models on structured + unstructured features with explainability, integrated as a real-time feedback loop during criteria authoring.
Trial completion risk	Factors including protocol amendments, disease indication, and criteria complexity predict whether a trial will complete or be suspended/terminated.	Survival analysis with competing risks (not just binary classification) with causal explanations.
Recruitment rate benchmarking	Current benchmarking fails to account for external factors like competing trials, SOC availability, and rare disease dynamics.	Simulation-based approach incorporating external competition, site characteristics, and indication-specific dynamics.

These challenges are well-characterized. What has been missing is a unifying architecture that treats them as components of a single intelligence system rather than as isolated ML problems.