FACTSPAN

www.factspan.com



Agentic AI Platform for Hypothesis Generation in Biopharma R&D - PoC

Executive Summary:

What if scientific hypotheses could be generated not just faster—but with traceable logic, real-time citations, and domain-specific reasoning? A global biopharma company exploring AI's role in early-stage R&D teamed up with Factspan to test this idea. The focus: apply Agentic AI to automate hypothesis generation for use cases like biomarker discovery and diagnostics.

Factspan designed a multi-agent prototype that orchestrated LLMs, integrated biomedical knowledge graphs, and produced audit-ready scientific outputs. Built for explainability and compliance, the early-stage platform showed how GenAI could complement human researchers in complex, high-stakes environments.

About the Client

A global biopharma company focused on innovation in oncology, immunology, and rare diseases. The organization operates multiple R&D hubs and upholds stringent standards around data security, regulatory compliance, and scientific validity.

Business Impact:

~60% reduction in literature review and drafting time

3x faster creation of structured assay reports

Complete traceability of outputs via embedded logging

Significantly reduced review effort through integrated human validation

Business Challenge

The client's R&D teams were in search of a digital co-pilot to accelerate early-stage scientific ideation. However, their existing research workflows were highly fragmented—navigating between structured databases, published literature, and clinical trial registries involved significant manual effort and slowed down hypothesis generation.

There was no unified way to reason across these diverse biomedical data sources, and the lack of traceable, explainable AI supported limited confidence in automation. The challenge was to design a secure, modular platform that could simulate scientific reasoning, maintain full auditability, and flexibly adapt to evolving research needs, all while meeting the compliance and data governance standards required in biomedical environments.

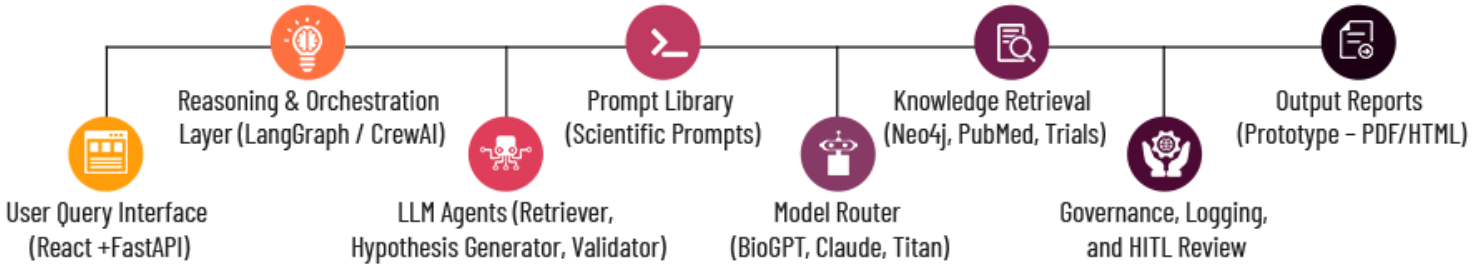
Our Solution

Factspan developed a focused Agentic AI system to simulate hypothesis generation workflows and assess GenAI's role in early-stage biomedical research.

Key components included:

- **Multi-Agent Orchestration:** Coordinated agents for literature review, hypothesis formulation, and assay mapping
- **Prompt Library:** Pre-configured and versioned scientific prompts tailored to biomedical reasoning
- **Orchestration Layer:** Used LangGraph and CrewAI to break down and manage reasoning flows
- **Knowledge Graph Integration:** Combined internal Neo4j graphs with public ontologies for grounded information retrieval
- **Model Routing:** Routed tasks to best-fit models including BioGPT, Claude, and Amazon Titan via Bedrock
- **Human-in-the-Loop Review:** Inserted checkpoints for expert review of AI-generated hypotheses
- **Traceability Layer:** Captured all decisions, prompt iterations, and model outputs for audit readiness

The system was evaluated in a secure, sandboxed environment using real but non-sensitive datasets and aligned with the client's infrastructure and data policies.



Note: This early-stage architecture was evaluated in a controlled, secure environment and tested on real-world research workflows using non-sensitive data.

Factspan is a Data and AI company. We help global enterprises and product companies to collect, store, and manage their varied and complex data structures, identify patterns and trends, and build analytical models to analyse and deliver business outcomes.