

From Pilot Purgatory to Production Intelligence: Scaling Agentic AI in Pharma

Sid Reddy, Barun Maskara



HUMAN + AI
i n t e l l i g e n c e

Speaker Intros



Sid Reddy

AI Innovation Leader, Humigent

- Founding member, **Google NotebookLM**
- Expert in LLMs, deep learning, RL & conversational AI
- AI roles at **Google, Amazon, Meta & Microsoft**
- 120+ publications & patents in AI
- BS CS IIT Kharagpur; PhD Arizona State



Barun Maskara

Founding Partner, Humigent

- Life Sciences strategist; **20+ years** in Life Sciences industry
- AI & advanced analytics for **pharma & biotech**
- Led teams at **ZS, Deloitte, PharmaACE, Boston Scientific**
- 70+ patents in Life Sciences
- BS EE IIT Kharagpur; MS BME Johns Hopkins; MBA Wharton



AGENDA

Today's discussion

- 01 The Problem** — why pilots stall
- 02 The Solution** — a path to production
- 03 Illustrative Example** — what it looks like in practice

THE PROBLEM

The Pilot Paradox: Pharma's AI Bottleneck

80%

**of pharma AI initiatives
launch as Proof of Concept**

30%

**successfully reach
production-grade deployment**

Why Pilots Fail: The Common Diagnosis

Organizational Reasons

- **Isolated Experiments**
- **Distrust**
- **Governance**

Common Misdiagnosis

Myth: We need more data

Reality: Data is siloed

Myth: We need a better model

Reality: LLMs are probabilistic

Myth: We need better prompts

Reality: Prompt engineering is a patch, not an architecture

Two Architectural Gaps Blocking Production

GAP 1 ► STRUCTURAL

- **General-purpose LLMs** cannot reliably reason over complex, siloed pharma data
- **Multi-table joins** across CRM, CDM, etc produce inconsistent results
- **Schema drift and column aliasing** cause silent failures that are impossible to detect downstream

Incorrect analytics delivered with high confidence

GAP 2 ► SAFETY

- **Probabilistic LLM outputs** cannot provide the deterministic guarantees
- **No native audit trail** for regulators and compliance teams
- **Hallucination risk** is non-zero by design

Every AI output becomes a compliance liability

Poll: Where Are You on the AI Journey?

Where does your organization sit today?



Attended demos, ran experiments. No formal business case.

Funded POC underway. Promising in sandbox. No production data yet.

Real users, inconsistent adoption. Debating invest vs. restart.

Processes real queries. Users trust it daily. Audit trail exists.

Multiple agents across domains. Self-improving. Handles exceptions.

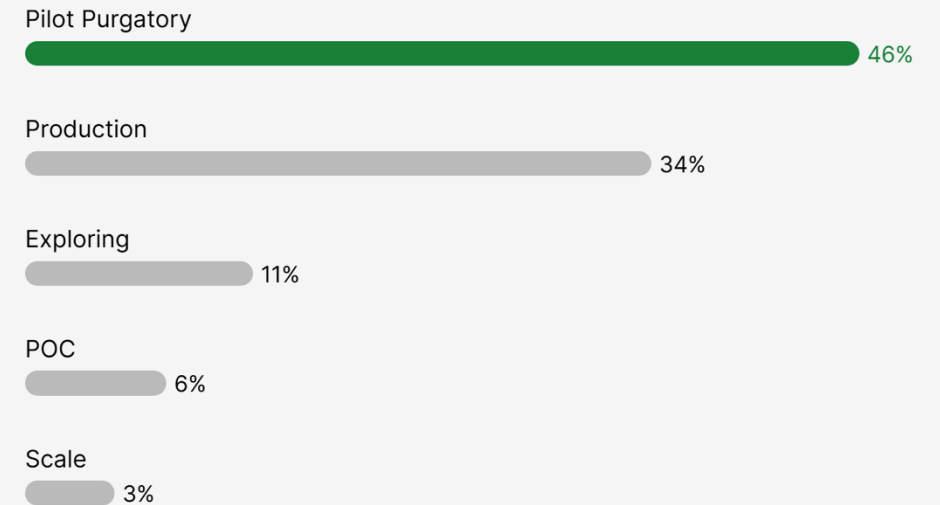


Join at
slido.com
#9632 412

Active poll

35

What stage is your organization currently in regarding AI implementation?



Part 2



Humigent

THE SOLUTION



Humigent Approach to Pharma AI

SOLUTIONS — AI-Powered

Data Strategy Data Lens	Analytics One Customer Universe	Analytics Dashboard Lens	Analytics Patient Explorer	+ More
Field Analytics IC Healthcheck	Forecasting GeoForecast	Medical Affairs MSL Voice	Market Research Insights Navigator	

ORCHESTRATOR

Routes, prioritizes, and sequences agents — one coordinating Decision Engine across all signals

AGENTS — Modular · Expandable

Ingestion Agent	Authentication Agent	Ask AI Agent	IC Reader Agent	+ More
Data Validation Agent	Authorisation Agent	Summarization Agent	Audit Agent	

INTELLIGENCE LAYER

R³ Cognitive Framework — Retrieve · Reason · Respond

RETRIEVE Pharma Data, Query-Ready Claims, formulary, EMR, Rx, field signals — unified and accessible	REASON Coordinated Agent Logic Controlled, reproducible, and explainable decision workflows	RESPOND Decisions, Not Just Answers Outputs tied to source and logic path
--------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------

FOUNDATION LAYER

HumigentLM — 30B parameter Pharma-Native Model

A domain-specific model trained on commercial, medical, and access data with Life Sciences vocabulary and context.

Commercial · Medical · Access Domain corpus	Pharma Vocabulary & Context Life Sciences-specific Ontology	Structured Data Understanding Pharma data reasoning
------------------------------------------------	----------------------------------------------------------------	--------------------------------------------------------

HumigentLM: A Pharma-Native Language Model

Built for regulated, high-stakes decisioning



PHARMA-NATIVE

- Trained on medical, commercial & market access corpus
- Understands pharma terminology natively



PRODUCTION-SAFE

- Runs inside client environment
- Privacy & security by design
- Compliance alignment built-in



PURPOSE-TUNED

- Fine-tuned for analytics, access & medical
- Not just an LLM wrapper
- Domain-specific from day one



COST-EFFICIENT

- Lower cost vs. frontier models
- Scales for operational workloads
- Improves for you with usage

✓ Benchmarked against GPT-5.4 and GPT-4.1 Mini for safety, accuracy, and cost

● Domain accuracy

● Privacy-first

● Self-improving

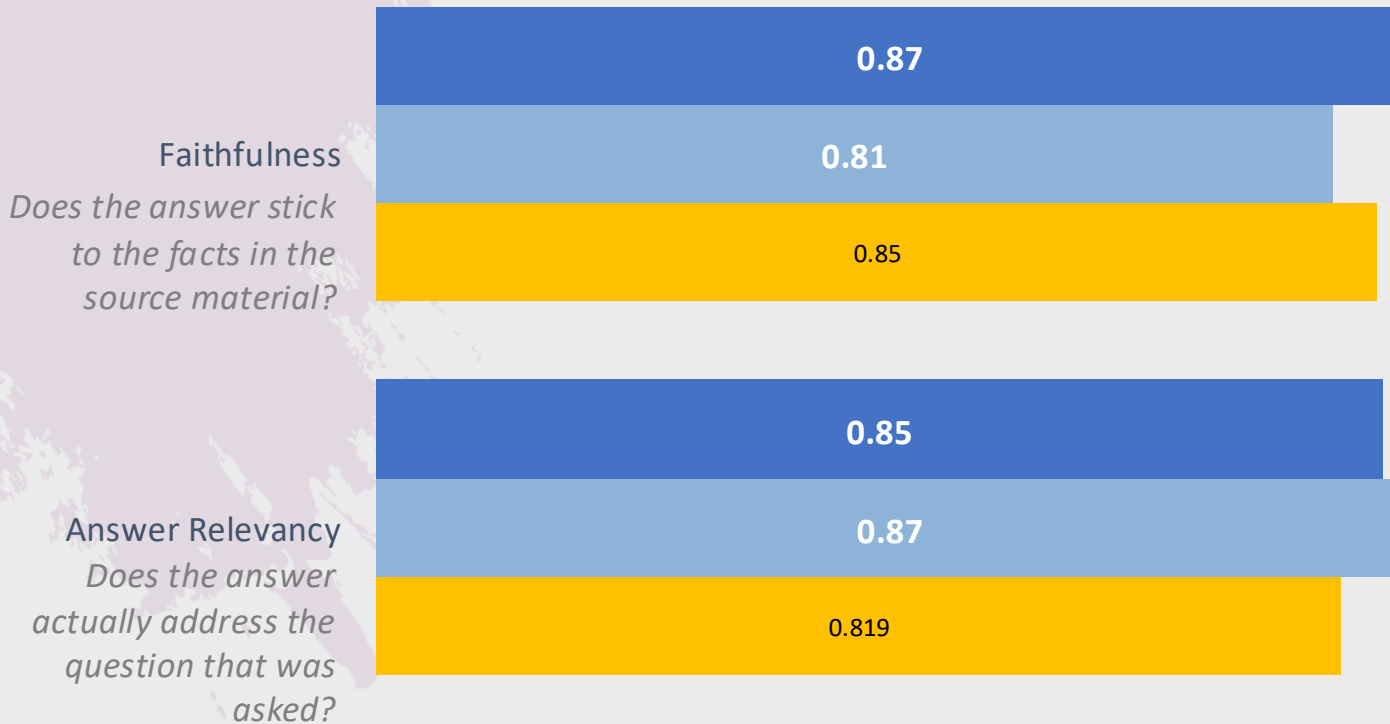
● Cost-efficient

Why it matters: Enables audited, production-grade AI—without sacrificing speed or cost.


HumigentLM vs. Competitors

Response Quality Scores

■ HumigentLM 0.4 ■ GPT-4.1 Mini ■ Claude Opus 4.6




* GPT-5.4 could not provide response that met the minimum threshold of response quality in many instances, leading to very low Response Quality Scores




Speed (seconds)

HumigentLM	16s ✓
GPT-4.1 Mini	92s
GPT-5.4*	24s
Claude Opus	228s



Compute Cost

HumigentLM	Low ✓
GPT-4.1 Mini	Medium
GPT-5.4*	High
Claude Opus	High



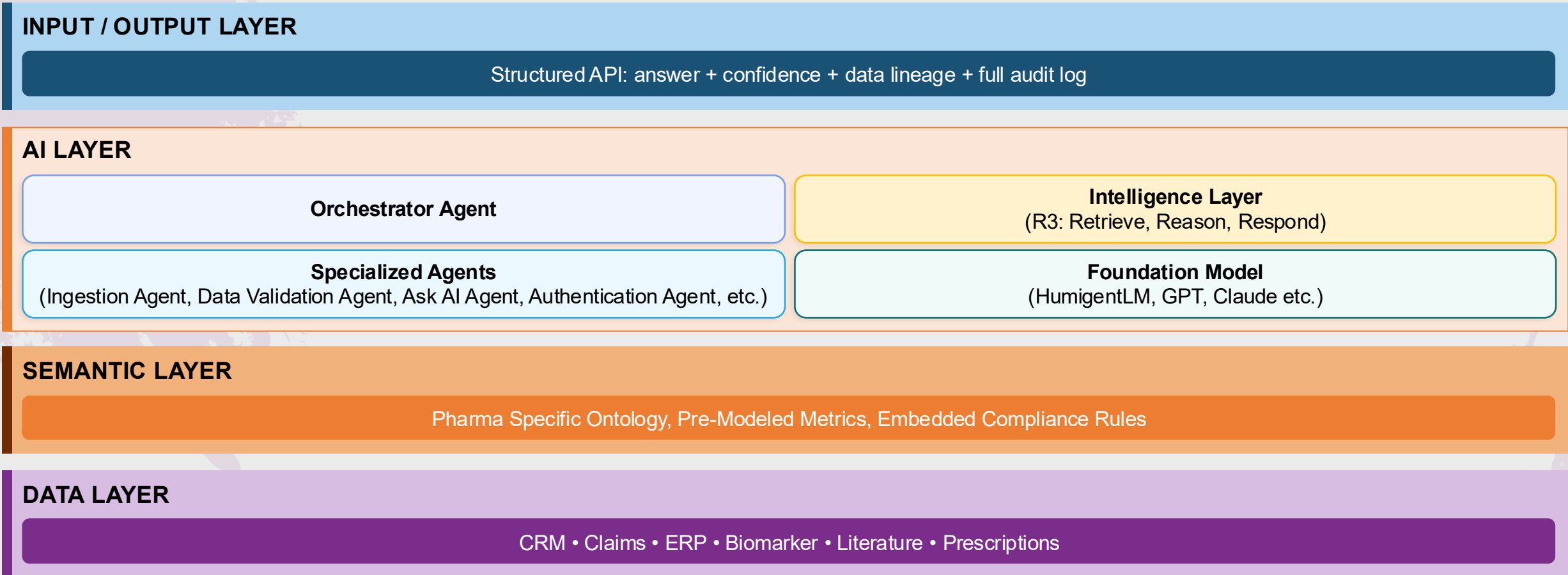
Privacy Controls

HumigentLM	High ✓
GPT 4.1 Mini	Medium
GPT-5.4*	Medium
Claude Opus	Medium

Full Solution Architecture: The Pharma Cognitive Engine



A layered stack where each tier has a defined role — from raw data to auditable output:





ILLUSTRATIVE EXAMPLE



The Text-to-SQL Challenge Setup

We will demonstrate a query that breaks standard LLMs — and show how the R³ system solves it correctly.

THE QUERY (Natural Language):

“Which of our top 50 HCPs by TRx volume saw a decline in new patient starts last quarter, despite an increase in rep call frequency?”

Why This Is Hard:



3-Table Join

CRM, Claims & Field Force Planning



Derived Metric

“New patient starts” requires business logic



Temporal Dependency

“Top 50 by TRx” is prior-quarter ranking



Cross-Table Deltas

QoQ changes across two separate fact tables

A standard LLM will produce a confident, plausible-looking SQL query that is factually incorrect.

Standard LLM vs. R³ Multi-Agent System

Standard LLM Response

- Joins HCP_ID ↔ Provider_ID — wrong keys, silent failure
- “New patient starts” computed as raw count — misses attribution rule
- Ranks by current TRx, not prior quarter — wrong cohort
- Returns formatted table with wrong data — no error signal

Result: Analyst presents incorrect HCP ranking to the leadership.

Audit trail: None. If challenged, cannot explain derivation.

R³ Multi-Agent System

- Schema Agent resolves NPI_ID from data catalog — alias resolved
- Reasoning Agent applies 90-day attribution window correctly
- Retrieval Agent ranks HCPs on prior-quarter TRx (Q3)
- Validation: non-negative, bounded, delta directions consistent

Result: Correct HCP ranking with full derivation trace.

Audit trail: Full proof attached. Compliance-ready.

Key Takeaways

Diagnose Before You Build

- AI pilots fail due to architectural gaps -- not data or model gaps.
- Structural Gap: LLMs cannot reason over siloed schemas.
- Safety Gap: Probabilistic models cannot guarantee auditable precision.
- Use the Pilot Diagnostic framework to audit data, logic, and governance gaps before investing further

Evaluate on Multi-Step Tasks

- Single-model accuracy is not the bottleneck — multi-step reasoning is
- R³ (Retrieval, Reasoning, Response) with specialized agents solves the reliability problem
- Full audit trails make outputs compliance-ready

THE BOTTOM LINE

The gap between demo and production is an architecture problem, not a model problem. Close it with architecture.

Please visit us at Humigent booth # 304 to continue this conversation



Thank You

Questions?