databricks

EXECUTIVE BRIEF

How Healthcare and Life Sciences Leaders Are Putting AI to Work

Based on Hundreds of Databricks Deployments

Introduction

Healthcare providers, health plans, biopharma and medtech organizations sit on enormous volumes of operational and scientific data, spread across electronic health record (EHR) systems, claims platforms, clinical trial databases, research pipelines and manufacturing systems. Data is often locked in PDFs, imaging files and unstructured notes, siloed across departments and functions. As a result, prior authorizations stall, chart abstraction backs up, HEDIS gaps stay open and trial protocol deviations go undetected until they impact patient outcomes and increase cost. Clinicians, care managers and researchers spend their days navigating workflows the platform should handle for them.

The companies pulling ahead aren't simply deploying AI; they're deploying AI on a unified, governed data platform. That's the difference between a pilot that impresses and a system that scales.

Key takeaways

Multi-agent architectures, not single-purpose chatbots, are the production standard for leading companies.

Novo Nordisk leveraged Databricks Agent Bricks and Genie to build their "Co-Scientist", which coordinates specialized agents to search across Novo's clinical data foundation and return curated, context-aware insights within seconds. With Databricks, they're driving $157M+ net new value attributed to optimized clinical trials.

The most effective AI works across every data type — structured and unstructured — in a single coordinated system.

Genie allows teams on the frontlines to query complex patient, trial and claims data in natural language, while Vector Search extracts structured signals from unstructured clinical notes and scientific documents. Lakebase supports FHIR-compliant patient engagement tools, real-time clinical trial monitoring and automated prior authorization applications.

Data governance is the architectural prerequisite for AI at scale.

From HIPAA and CMS to FDA and GxP requirements, governance is non-negotiable in healthcare and life sciences. Unity Catalog's fine-grained access controls, PHI protection and full lineage through agent workflows allow organizations to scale AI across clinical, claims and research data without introducing compliance risk.

Time to value is measured in weeks, not quarters.

Across Databricks deployments, companies are seeing results faster than expected. Biogen took a data pipeline that previously required 2 weeks to process 700,000 genomic variants, and optimized it to annotate 2 million variants in just 15 minutes. Real Chemistry quickly scaled their AI innovation, rolling out various custom-built POCs in just one week with Databricks Apps.

The efficiency gains are real and attributable.

After optimizing their clinical trials, Novo Nordisk improved operational efficiency by $14M. Walgreens experienced a 20% increase in demand planning productivity.

Align your stakeholders before you build.

The most successful deployments involve CMIOs, revenue cycle leaders, Chief Data and Analytics Officers, heads of clinical development, sales leaders, IT security and finance from day one. This cross-functional collaboration is what separates pilots that scale from pilots that stall.

2.5x	Faster Clinical Abstraction
$14M	Operational Efficiency Gains
5–7x	Faster Payer Insights

Use cases

Multi-Agent Supervisor Architectures

Healthcare and life sciences organizations rely on dozens of systems across EHRs, claims platforms, clinical trial systems and research environments. Multi-agent architectures solve this by assigning specialized agents to handle different tasks — from prior authorization and clinical documentation to trial monitoring and research synthesis — all coordinated by a central supervisor that routes each request to the right system.

AstraZeneca built a multi-agent system on Databricks in just 60 minutes that parsed over 400,000 clinical trial documents, turning complex, unstructured clinical documents into structured, analytics-ready evidence in a single governed workflow. The same architectural pattern powers Novo Nordisk's AI "Co-Scientist", which coordinates specialized agents across Agent Bricks and Genie to surface context-aware insights on adverse events, patient subpopulations and trial protocols in seconds, replacing weeks of manual effort.

Built on: *Unity Catalog (access controls and data lineage), Genie agents (structured data/SQL), RAG agents (unstructured documents), function-calling agents (vector search), Databricks Model Serving*

Natural Language Data Access and Analytics

What if clinicians, care managers, researchers and analysts could get answers from complex clinical, claims and research data just by asking a question — no SQL, no analyst queue, no waiting? With Genie, healthcare and life sciences teams can query patient records, trial data and operational metrics in natural language, reducing time to insight from days to minutes and enabling faster, data-driven decisions at the point of care. Care managers get direct access to HEDIS measure performance while maintaining strict data lineage for regulatory audit.

Milliman MedInsight uses Databricks to deliver risk and utilization insights to health plans, providers and public agencies, reporting 15–35x improved workload performance and a 5–7x increase in the speed of insight delivery to customers.

Built on: *Databricks Model Serving, open-source LLM support (Llama, Mistral), Genie text-to-SQL, Mosaic AI critique agent framework*

Predictive Patient Outcomes

Predictive models use real-time data from EHRs, claims, devices and clinical systems to identify risks before they become adverse events. From patient deterioration and readmissions to trial delays and drug manufacturing failures, reactive decision-making is one of the most costly and outcome-critical problems in healthcare.

On Databricks, teams can train and deploy ML models on unified patient data to identify high-risk patients before adverse events occur. Flo Health uses machine learning to deliver personalized health predictions and insights to millions of users, improving engagement and enabling earlier, data-driven interventions at scale.

Children's National's AI/ML team leverages Unity Catalog for data and model governance, ensuring consistent and secure collaboration across teams, so they can scale AI initiatives with confidence and directly improve patient care outcomes, with model training timelines reduced from months to minutes.

CVS Health uses Databricks to power a personalization engine that tailors patient outreach across the pharmacy footprint, lifting medication adherence by 1.6%, a meaningful gain considering non-adherence costs the US healthcare system an estimated $528B annually.

Built on: *Databricks streaming pipelines, Delta Lake, real-time ML model serving, Mosaic AI*

Data Governance as Enabler

Governance is what makes enterprise-wide AI deployment possible by giving organizations the confidence to scale. Unity Catalog's access controls, data masking and lineage tracking are built into the platform, so AI agents can operate across sensitive patient records and clinical trial datasets without creating compliance risk. This ensures compliance with HIPAA, CMS, FDA and GxP requirements, with full auditability across patient and research data.

For example, Flo Health uses Unity Catalog to enforce fine-grained access controls and data governance across sensitive health data, enabling award-winning features like Anonymous Mode while maintaining compliance at scale. Nearly half of its employees now use Databricks each month to securely access data and drive AI-powered insights.

Built on: *Unity Catalog (column-level security, dynamic data masking, data lineage tracking through agent workflows, audit logging), fine-grained access controls for multi-agent systems*

Strategic Recommendations

Build Your Foundation

Before your first agent deployment, implement Unity Catalog. Its built-in access controls, data lineage and audit logging are what allow you to expand AI beyond your first team without creating governance debt. Start by auditing high-volume, manual workflows across clinical, revenue cycle and research operations (prior authorization, clinical documentation, care gap analysis, clinical data abstraction, etc.) and identify the highest-impact opportunities for automation. Then, run a focused pilot on one real use case to prove value quickly and establish a repeatable foundation for scale.

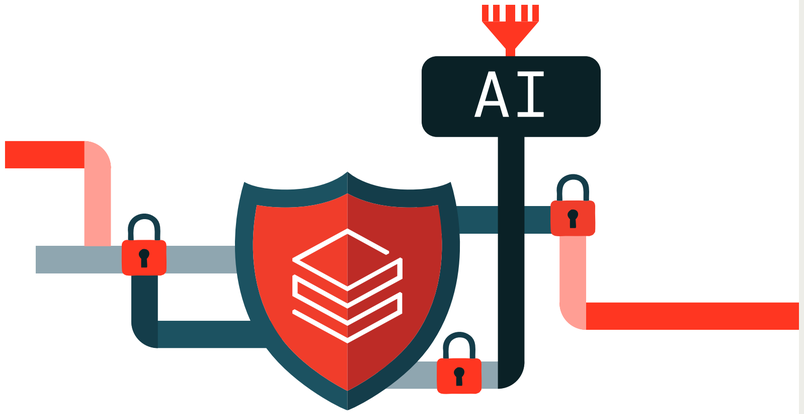
Scale What Works

Once a workflow is proven in one business unit, treat it as a template for the next. Databricks enables reusable pipelines, shared notebooks and governed data assets that let you take a text-to-SQL workflow or multi-agent architecture and expand it across an organization without rearchitecting. Use MLflow to track model performance and Unity Catalog to manage access as new teams and data domains come online. The goal is to build a library of reusable agent components instead of a collection of isolated deployments.

Avoid These Traps

Not every investment is worth making. Three to avoid:

- **DIY MLOps infrastructure.** Databricks and major cloud providers have already solved this; building your own means spending engineering time on maintaining infrastructure instead of solving business problems.
- **Single-purpose AI point tools.** Every additional vendor increases the total cost of ownership and adds integration debt.
- **Premature model optimization.** LLMs improve every quarter. Build on an open architecture that lets you swap in better models as they become available without rebuilding your workflows.



Results from the Field

Three patterns emerge across these deployments:

Efficiency gains are concentrated in data operations: processing, querying and enrichment. This repetitive, high-volume work is exactly what AI handles best.

Speed improvements are transformational, not incremental. As one of the largest pharmacy networks in the US, Walgreens now processes 40,000 data events per second on Databricks, applying those insights to fill 825 million prescriptions a year with 20% higher demand-planning productivity and millions in supply-cost savings, directly strengthening patient access and affordability.

Scale compounds as each AI workflow absorbs more throughput without adding operational cost. Biogen isn't handling 2 million genomic variants with more people; the platform handles volume that would be impossible to manage manually.

In every case, the underlying enabler is the same: Databricks Platform that connects raw operational data to the AI layer without requiring teams to stitch together point solutions to bridge the gap.

Biogen	99.9%+	Faster Pipeline Processing 2 weeks → 15 minutes
novo nordisk®	$157M	Net New Value From optimized clinical trials
CVSHealth.	1.6%	Adherence Lift across medications
AstraZeneca	60min	Week Dev Time Prototypes → Working POCs

Realistic timeline

When it comes to AI, business decision-makers often wonder: How long before this technology delivers something real?

Based on Databricks deployments across healthcare and life sciences organizations:

1 WEEK

Development Time

Real Chemistry can now launch working POCs within a week, enabling rapid iteration or pivoting as needed

MINUTES

Model Training

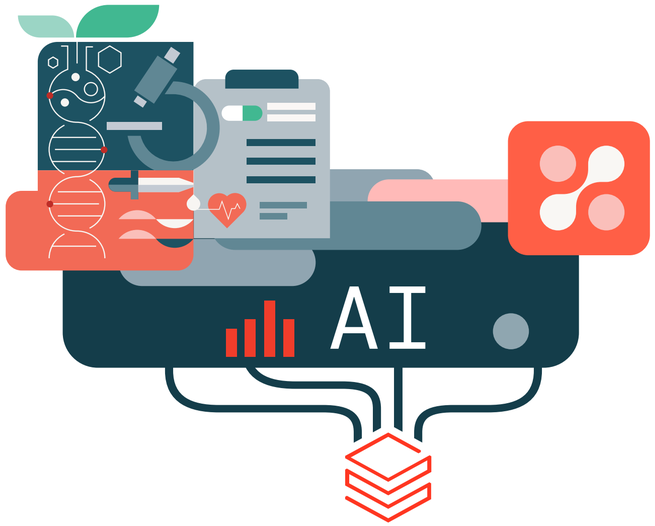
Children's National Medical Center delivers better pediatric outcomes with model training timelines reduced from months to minutes

9 MONTHS

Enterprise-Scale Rollout

Novo Nordisk went from initial pilot to a governed platform used by over 20% of its clinical trial development organization

Real, measurable value can be delivered in weeks, not the multi-year transformation cycles most enterprises are used to. The companies that move fastest have two things in place before they start: a governed data foundation, so agents have clean, accessible data to work with, and a clear business owner who defines what 'working' looks like. Both are achievable before your first sprint begins.



Conclusion

Healthcare and life sciences AI is no longer an experiment; it's an operational capability that leading companies are deploying at scale right now. The companies pulling ahead aren't doing it with a collection of point solutions or a single chatbot. They're doing it with a unified platform that brings data engineering, governed access, multi-agent orchestration and model serving together in one environment.

Databricks is the defensible choice for healthcare and life sciences AI at scale, and it comes down to three advantages:

Open architecture:
Native support for ChatGPT, Claude, Llama and other open-source models means no model lock-in, and you never have to rebuild when a better model emerges.

Governance at the platform layer:
Unity Catalog makes it possible to deploy AI broadly across an organization without compromising data protection or compliance.

The compounding advantage of a unified platform:
Every new use case built on Databricks shares the same data assets, pipelines and governance framework, so the tenth deployment takes a fraction of the effort the first one did.

The question for industry leaders isn't whether to deploy AI. It's whether to spend the next two years stitching together tools that don't scale, or build on a unified foundation where every deployment makes the next one faster, smarter and more efficient.

databricks