



EXECUTIVE E-BOOK · LIFE SCIENCES

Accelerate your data and AI transformation

A molecule-to-market playbook for pharma and biotech function leaders,
built on the Databricks Data Intelligence Platform.

The data is already there. The **value** is still waiting.

Few industries generate as much data as pharma and biotech: omics and imaging in discovery, EDC and CTMS feeds in clinical, historian and batch data on the plant floor, CRM, claims and engagement data in commercial. Yet most of it still lives in function-specific systems, under different rules, owned by different teams. Insight moves slower than the science.

Generative AI has raised the stakes. Every function now has credible use cases, from drafting clinical documents to investigating manufacturing deviations. But each of them depends on the same thing: trusted, governed, connected data.

This e-book is written for the leaders who own those outcomes: heads of R&D, clinical operations, manufacturing and supply, commercial, medical affairs, regulatory, safety, finance and HR. It follows the value chain from molecule to market and shows, function by function, where value sits, which use cases to start with, and how a single data and AI platform makes them work together.

Inside

01 One value chain, one foundation · 02 Research & discovery · 03 Clinical development · 04 Manufacturing & supply · 05 Commercial & medical · 06 Regulatory, safety & enterprise · 07 GxP and responsible AI · 08 Prioritize and get started

\$60–110B

Potential annual economic value of generative AI for pharma and medical products¹

~\$500K

Estimated lost sales for each day a drug's development is delayed²

~60%

Of life sciences C-suite executives plan to increase gen AI investment across the value chain³

\$2.6B

Estimated cost to bring one new approved drug to market, including the cost of failures⁴

1 McKinsey, "Generative AI in the pharmaceutical industry: Moving from hype to reality," Jan 2024. 2 Tufts CSDD, as reported in Applied Clinical Trials, June 2024 (2023 USD). 3 Deloitte, 2025 Life Sciences Executive Outlook (survey of 150 executives). 4 Tufts CSDD, 2014 (2013 USD; includes time and failure costs).



One value chain. One **data foundation**.

Most AI programs stall because each function builds its own stack. Leaders who move fastest treat data as a shared asset that flows from molecule to market, governed once and used everywhere.



PRINCIPLE 1

Start with outcomes, not tools

Anchor every initiative to a function KPI: cycle time, right-first-time, launch uptake. If a use case cannot name its metric, it waits.

PRINCIPLE 2

Govern once, use everywhere

One catalog, one access model and one lineage trail across data, models and agents. It is what makes GxP, privacy and partner sharing manageable at scale.

PRINCIPLE 3

Crawl, walk, run

Prove value in one or two functions, reuse the data products you built, then expand. Each wave should be faster and cheaper than the last.



Research & discovery

Turn fragmented scientific data into a reusable engine for target and molecule decisions.

\$15–28B

Estimated annual gen AI value in research and early discovery¹

The challenge

Scientific data is spread across ELNs, LIMS, instruments, public databases and partner CROs. Multi-omics and imaging volumes outgrow legacy systems, and scientists spend too much time finding and cleaning data instead of using it. Little of it is truly FAIR (findable, accessible, interoperable, reusable).

WHERE TO START

DATA ENGINEERING

FAIR research data products

Ingest ELN, LIMS and instrument outputs into curated, catalogued tables that every program can reuse.

MACHINE LEARNING

Multi-omics target identification

Integrate genomic, proteomic and phenotypic data to rank targets and biomarkers with explainable models.

GENAI AGENT

Scientific evidence assistant

An agent that searches and summarizes literature, patents and internal reports, citing its sources for scientist review.

ML + GENAI

Predictive chemistry

Property and ADMET prediction models that shorten design-make-test-analyze cycles.

DATABRICKS BUILDING BLOCKS

- Lakeflow to ingest instrument, ELN and LIMS data
- Unity Catalog for lineage and access across internal and partner data
- Agent Bricks and MLflow to build, evaluate and track models and agents

MEASURE SUCCESS

- Time from hypothesis to validated target
- % of research data catalogued and reusable
- DMTA cycle time per program



Clinical development

Use data and AI to design better trials, enroll faster and get to database lock sooner.

\$13–25B

Estimated annual gen AI value in clinical development¹. Every day of delay costs roughly \$40K in direct trial costs and ~\$500K in lost sales².

The challenge

Trial data is split across EDC, CTMS, eTMF, central labs, wearables and multiple CROs. Protocols keep growing in complexity, enrollment slips, and teams spend weeks authoring and reviewing documents by hand.

WHERE TO START

MACHINE LEARNING

Site selection and enrollment forecasting

Combine historical site performance with real-world data to pick sites and predict enrollment risk early.

GENAI AGENT

Clinical document authoring

Draft protocol sections, patient narratives and CSR content from source data, with human review built in.

ANALYTICS + ML

Risk-based quality management

Central monitoring that flags site and data anomalies before they become findings.

ANALYTICS

Real-world evidence

Governed, de-identified RWD for feasibility, external comparators and post-approval evidence.

DATABRICKS BUILDING BLOCKS

- Delta Sharing to exchange data with CROs and vendors without extra copies
- Unity Catalog fine-grained access and masking for patient data
- Genie so study teams can ask trial-ops questions in plain language

MEASURE SUCCESS

- Enrollment rate vs. plan
- Last-patient-out to database lock
- Document authoring cycle time



Manufacturing & supply chain

Connect plant, quality and supply data to make every batch right the first time.

\$4–7B

Estimated annual gen AI value across sourcing, manufacturing, quality and supply chain¹

The challenge

Operational data sits in MES, historians, LIMS, QMS and ERP, often split between OT and IT teams. Deviation investigations rely on manual searches through past records and SOPs, and supply disruptions are spotted late.

WHERE TO START

MACHINE LEARNING

Golden batch and yield analytics

Model process parameters from historian and MES data to predict yield and catch drift in real time.

GENAI AGENT

Deviation and CAPA copilot

Searches past deviations, SOPs and batch records to suggest likely root causes for the investigator to confirm.

MACHINE LEARNING

Predictive maintenance

Sensor-driven models that schedule maintenance before critical equipment fails.

ML + ANALYTICS

Supply resilience

Demand sensing, inventory optimization and supplier-risk views across the network.

DATABRICKS BUILDING BLOCKS

- Lakeflow streaming from historians and IoT sensors
- Unity Catalog lineage and audit trails that support ALCOA+
- Agent Bricks for copilots grounded in your SOPs

MEASURE SUCCESS

- Right-first-time rate
- Deviation closure time
- OEE and on-time-in-full



Commercial & medical affairs

Reach the right HCPs and patients with the right message, and turn field insight into action.

\$21–35B

Estimated annual gen AI value in commercial (\$18–30B) and medical affairs (\$3–5B)¹

The challenge

HCP, claims, CRM, digital engagement and third-party data rarely come together in one governed view. Content takes weeks to clear medical, legal and regulatory (MLR) review, and valuable insights stay buried in MSL notes and medical information requests.

WHERE TO START

MACHINE LEARNING

HCP 360 and next best action

A governed customer view that powers channel, timing and message recommendations for field and digital teams.

GENAI

Modular content at speed

Generate on-label content variants with claim tagging that makes MLR review faster, not optional.

GENAI AGENT

Medical insights mining

Summarize themes from MSL notes, advisory boards and medical information requests for strategy teams.

SELF-SERVICE ANALYTICS

Brand and field analytics

Brand, sales ops and market access teams ask questions in plain language and get governed answers.

DATABRICKS BUILDING BLOCKS

- Unity Catalog to enforce consent and third-party data licence terms
- Clean Rooms and Delta Sharing for privacy-safe partner collaboration
- Genie for self-service insight

MEASURE SUCCESS

- Launch uptake vs. forecast
- Content concept-to-approval time
- Field adoption of recommendations



Regulatory, safety & enterprise functions

These teams touch every product at every stage. They are often the biggest beneficiaries of a shared foundation because they spend the most time pulling data from everyone else's systems.

REGULATORY AFFAIRS

From assembling to authoring

Submissions draw on data from every function, often reconciled by hand.

- **GenAI:** agents that draft and QC submission sections from source data
- Regulatory intelligence monitoring across health authorities
- Traceable links between submitted content and source data

KPI: submission preparation time, health authority query turnaround

PHARMACOVIGILANCE

Scale safety without scaling headcount

Case volumes and data sources keep growing; the timelines don't move.

- **GenAI:** case intake, triage and literature screening with human review
- Signal detection across spontaneous reports, RWD and social data
- Unified safety data for aggregate reports

KPI: case processing time, on-time reporting rate

FINANCE, HR & PROCUREMENT

Run the business on the same data

Planning still depends on ERP extracts and spreadsheets.

- R&D portfolio, spend and forecast analytics
- Workforce and skills analytics to plan AI upskilling
- **GenAI:** Genie self-service for FP&A and procurement teams

KPI: forecast accuracy, close cycle time, time to insight

Why it compounds

The data products built for clinical, manufacturing and commercial use cases become the inputs regulatory, safety and finance teams need. Each new use case gets cheaper because the governed foundation already exists.



Compliant by design: **GxP** and responsible AI

In life sciences, an AI use case is only as valuable as your ability to defend it to an auditor. Build compliance into the platform once, rather than into every project.

Data integrity

Meet ALCOA+ expectations with end-to-end lineage, audit trails and versioned data.

Supports 21 CFR Part 11 and EU Annex 11 controls

Risk-based validation

Validate against intended use following GAMP 5 and computer software assurance principles.

Keep validated GxP zones separate from exploration areas.

Validate what matters, document once

Privacy and access

Row- and column-level security, masking and attribute-based policies for patient and HCP data.

HIPAA, GDPR and consent requirements

Responsible AI

Human-in-the-loop for regulated decisions, systematic evaluation, monitoring for drift, and guardrails on model endpoints.

Traceable outputs for FDA and EMA scrutiny of AI

HOW THE PLATFORM HELPS

- **Unity Catalog** governs tables, files, models and agents with one permission model and automatic lineage
- **MLflow** tracks every model and agent version, its training data and its evaluation results
- **Delta Lake** time travel keeps a reproducible record of data as it was used

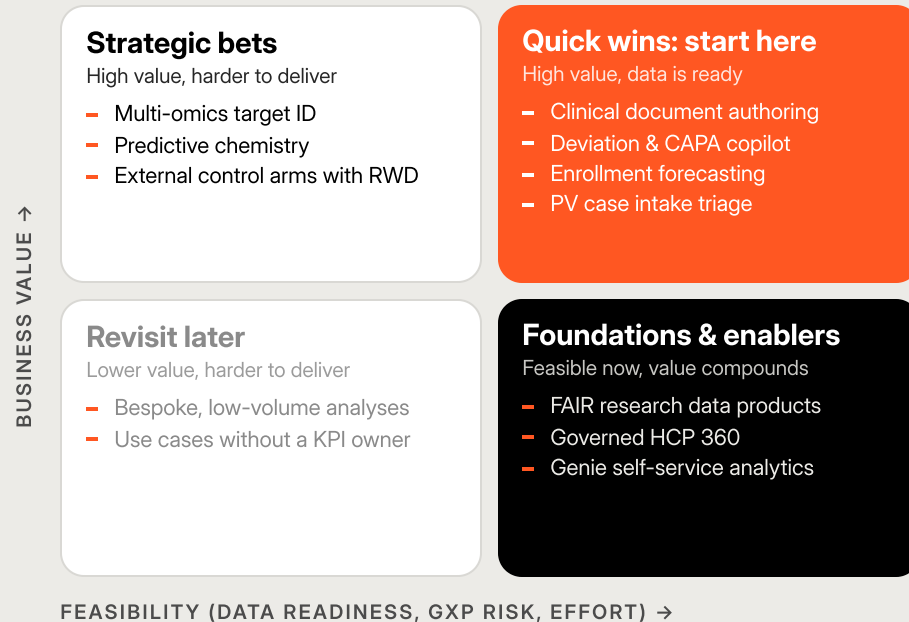
QUESTIONS EVERY FUNCTION HEAD SHOULD ASK

- Can we show which data trained this model, and who approved it?
- Where does a human review AI output before it reaches a regulator, patient or HCP?
- Can we share data with a partner without losing control of it?



Prioritize, prove, then **scale**

SCORE USE CASES ON VALUE AND FEASIBILITY



A 90-DAY PLAN

Days
0–30
Align

Inventory use cases with function heads, score them, assess data readiness and agree target KPIs.

Outputs: use-case portfolio · KPI baselines · data readiness view

Days
31–60
Build the foundation

Stand up governance in Unity Catalog, land priority data, and build the first one or two use cases with real users.

Outputs: governance live · priority data products · 1–2 pilots

Days
61–90
Prove and scale

Move to production with the right validation, measure against baseline, train champions, and queue the next wave.

Outputs: validated release · value vs. baseline · wave-two roadmap

NEXT STEP

Book a Data & AI Strategy Workshop with Kasmo Global

Bring your function leaders. Leave with a prioritized use-case map across your value chain, a data readiness view and a 90-day plan on Databricks.

Kasmo Global · Life sciences data & AI

[Book your workshop →](#)

 Email marketing@kasmoglobal.com
kasmoglobal.com
