

# CDER ONE DATA LAKEHOUSE FOR EFFICIENT AI & DATA GOVERNANCE

## **CDER Office of Business Informatics:**

Ethan Chen, CDER Director Division of Data Management

Venu Boppana, CDER Strategy & Innovation Leader

## **Databricks AI Summit**

June 10-13, 2024



The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance on behalf of the U.S. Food and Drug Administration



# AGENDA

- 1** FDA CDER at a Glance
- 2** What is CDEROne?
- 3** CDEROne AI
- 4** Moving Forward

CDEROne Analytics:  
**FDA CDER at a Glance**



# FDA at a Glance



20%

of U.S. consumer spending is  
on **products that FDA  
regulates**

FOOD

DRUGS

Over-the-counter and prescription drugs,  
including biological therapeutics and generic  
drugs

BIOLOGICAL PRODUCTS

MEDICAL DEVICES

COSMETICS

TOBACCO

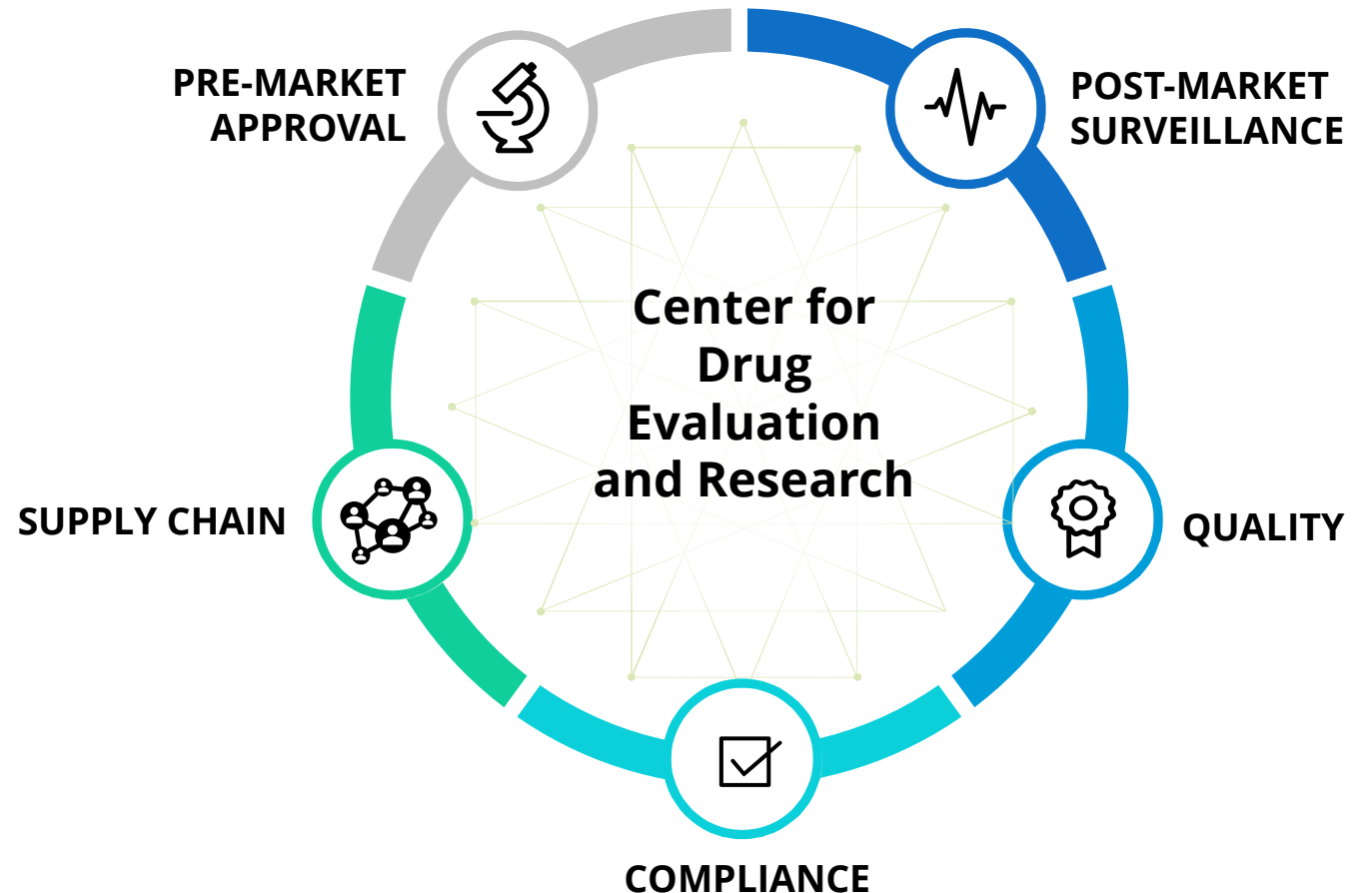
ANIMAL HEALTH

INSPECTION & ENFORCEMENT



# CDER's Mission

*...is to protect and promote public health by helping to ensure that human drugs are **safe and effective** for their intended use, that they **meet established quality standards**, and that they are **available to patients**.*





# Advancing CDER's AI Maturity

## DIRECTIVES

*Executive Order on AI, PDUFA, GDUFA commitments*

## INDUSTRY TREND

*CDER must modernize to keep pace with rapidly advancing industry trends in AI space*

## CDER STRATEGIC GOAL

*Commissioner Califf has emphasized the value AI can provide to CDER's mission*

[The Executive Order] establishes a **government-wide effort** to guide **responsible AI development and deployment** through federal agency leadership, regulation of industry, and engagement with international partners...It includes **8 requirements for HHS**, with specific language related to drugs and devices regulations.

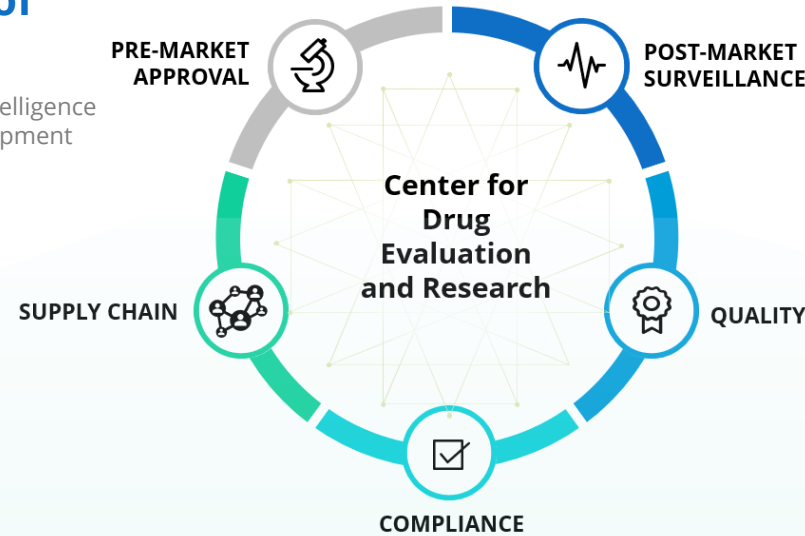
- EO 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

# Growing Challenges

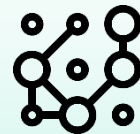
“An analysis of regulatory submissions of drug and biological products to the US Food and Drug Administration from 2016 to 2021 demonstrated an **increasing number of submissions that includes AI/ML**”

- Liu, Q. et .al Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2022. *Clin. Pharmacol. Ther.* 113(4), 771-774

Challenges that impacted across business areas necessitated **improved data & analytics modernization at an enterprise level** to replace the legacy siloed on-prem systems



Increased Number of Submissions



Growing Regulatory Complexity



Siloed Data



Need for self service capabilities



Need for AI capabilities

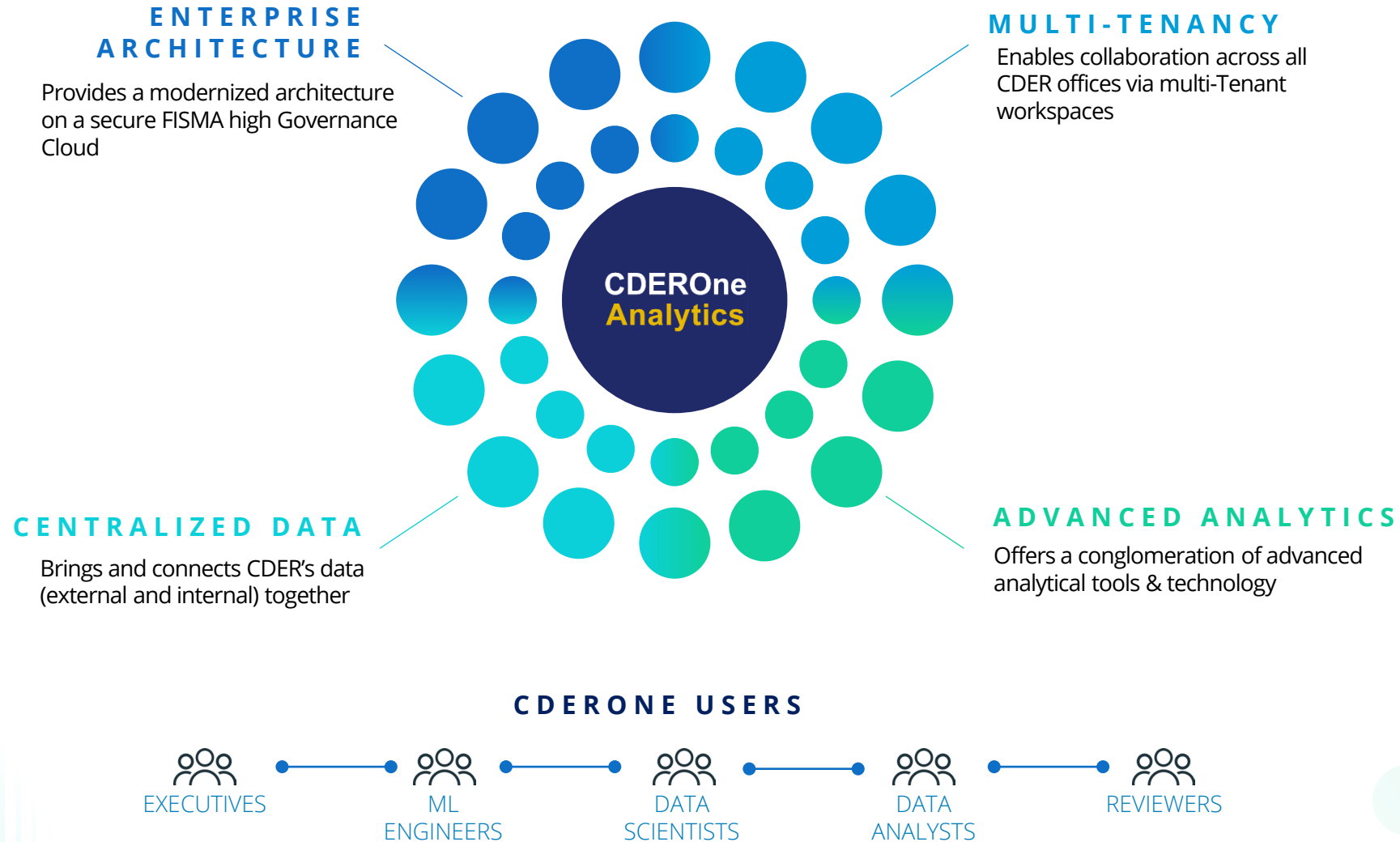


# CDEROne Analytics: What is CDEROne?



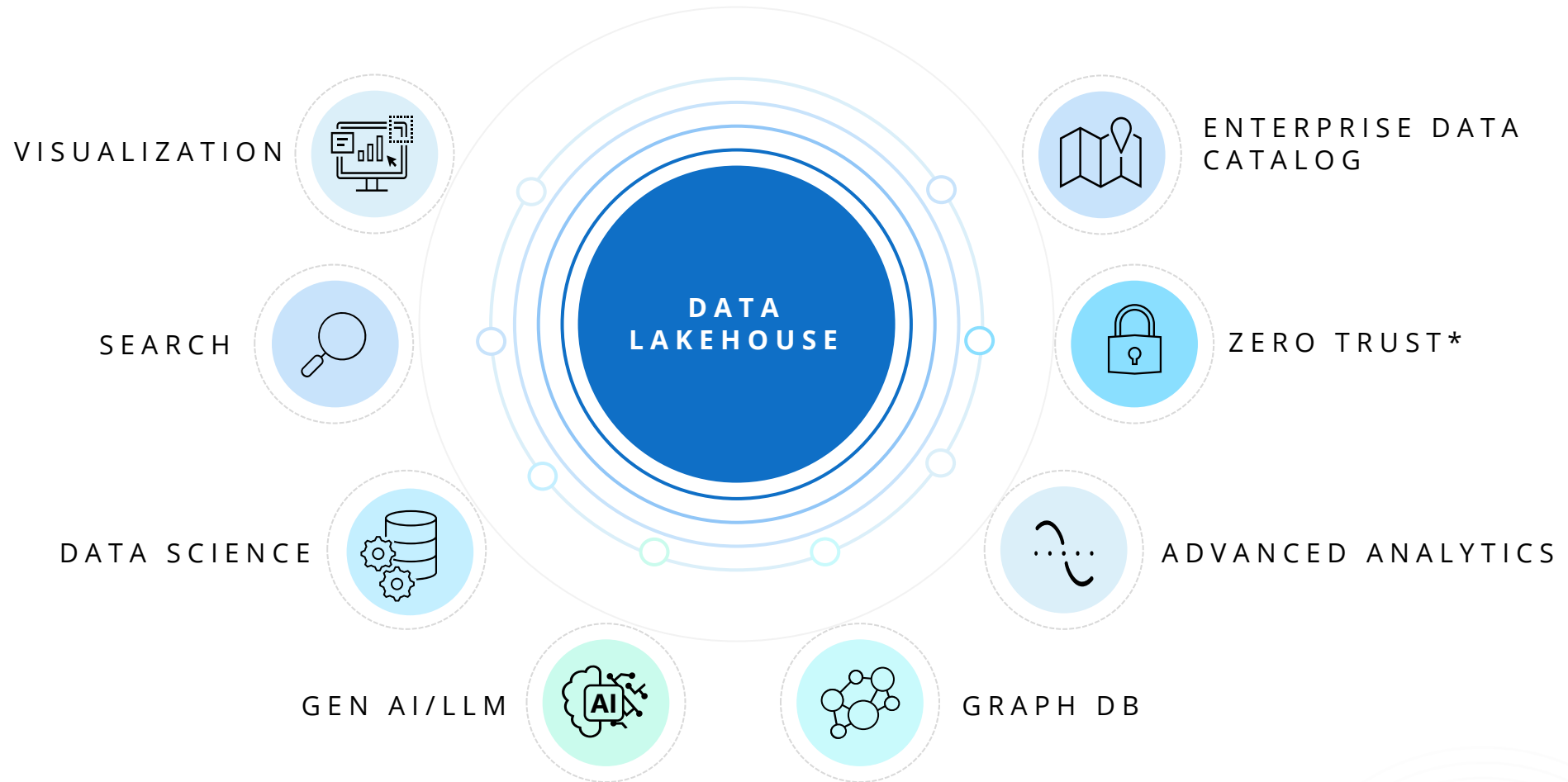
# Why CDEROne

CDEROne Analytics is a secure Enterprise Platform that acts as a centralized one-stop-shop, integrating several data sources to provide tailored, self-service, augmented analytical solutions that support the business needs of the Center for Drug Evaluation and Research (CDER).



# CDEROne Analytics Capabilities

CDEROne is equipped with industry leading data analytics capabilities, including Databricks



\*In-Progress, ETA: 2024 1<sup>st</sup> Quarter

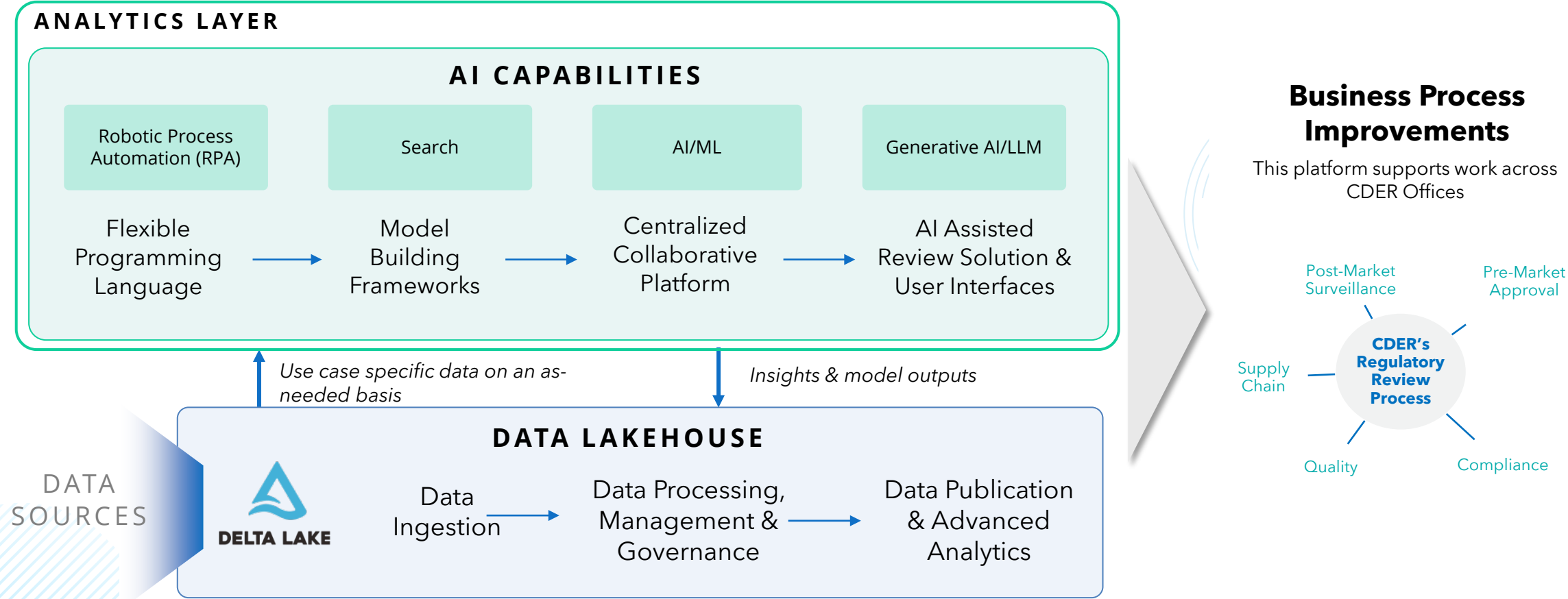


CDEROne Analytics:  
**CDEROne Data and AI**



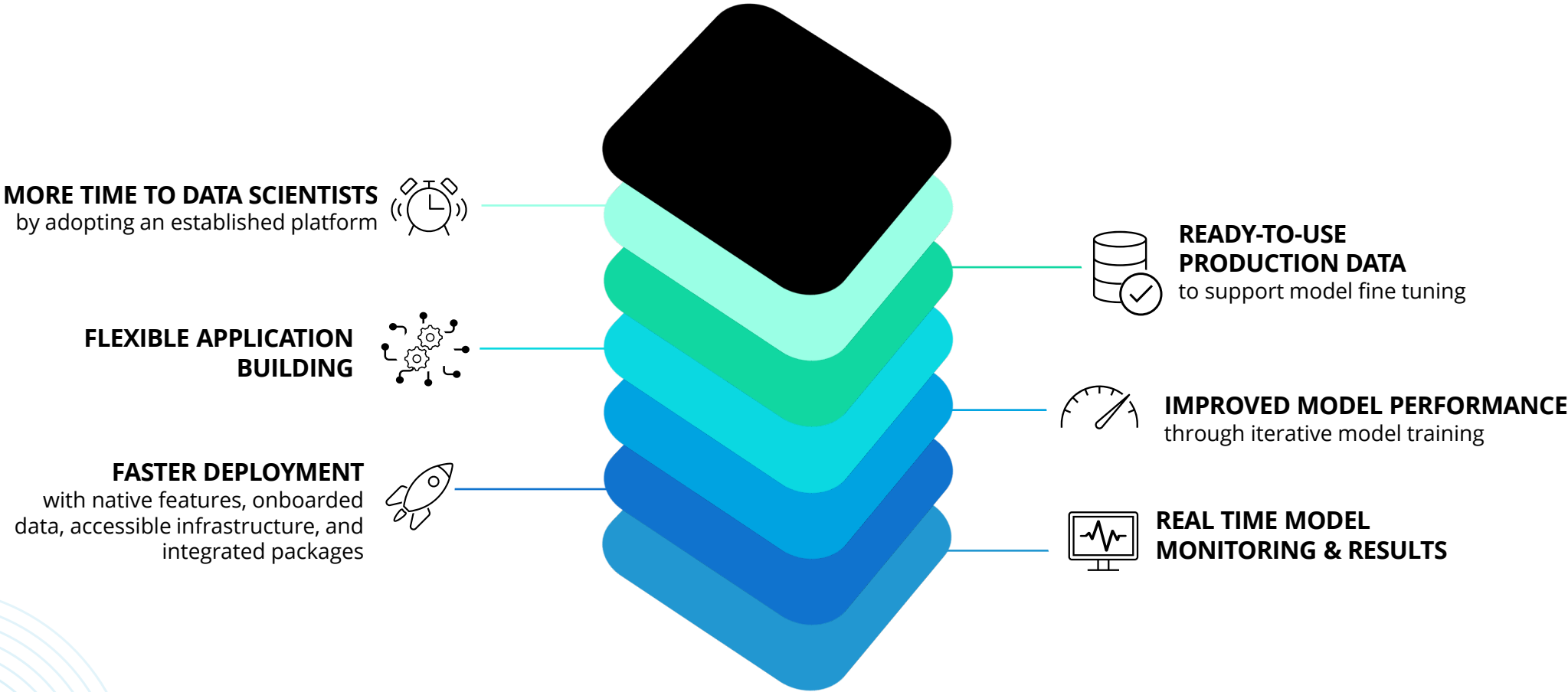
# CDEROne's Data Lakehouse and Analytics Layer

The AI capabilities, built on top of the data layer, encompasses the tools and technologies that are made available to CDER data scientists.



# Platform Benefits to Data Scientists

Data Scientists build more finely tuned models improving performance, offload compute of model training freeing up time, and develop flexible ways to deliver model results directly to business stakeholders.



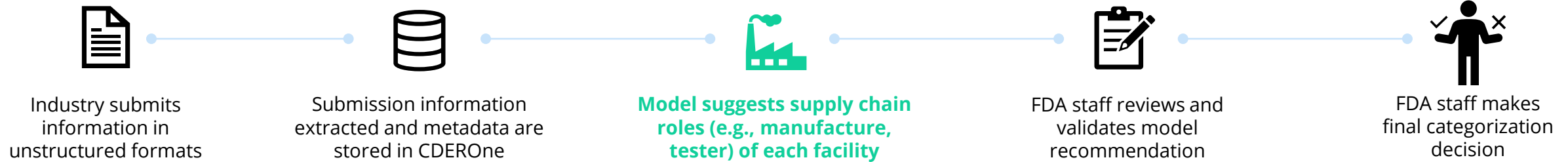
DATABRICKS NATIVE & BUILT-IN CAPABILITIES





# USE CASE 1: Manual Processing of Unstructured Content Supported by ML

Leveraged AI to support the manual process of classifying industry submitted information (e.g., facility business operation).



## DATA LAKEHOUSE

End-to-end data pipeline with centralized data engineering and analytics tools and technologies



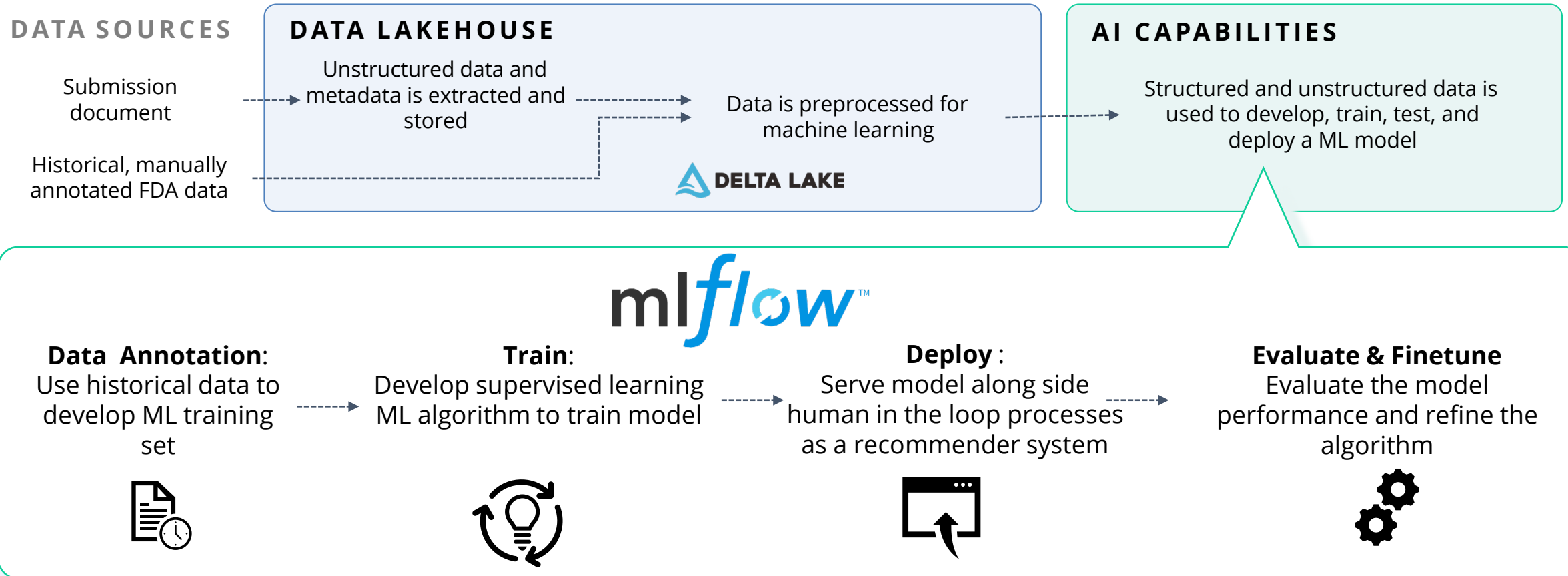
## AI CAPABILITIES

Advanced analytics tools and technologies made available to CDER data scientists.

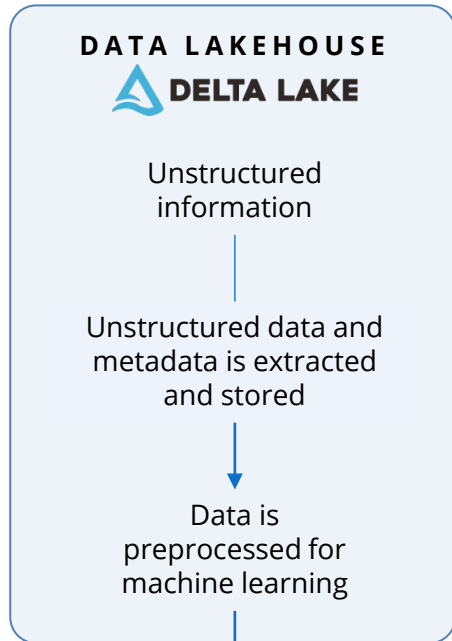


## BUSINESS PROCESS IMPROVEMENTS

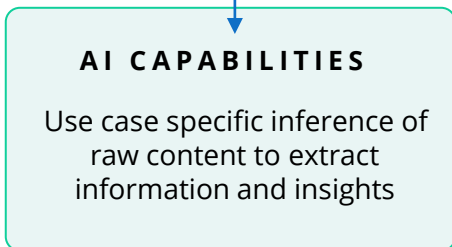
# USE CASE 1: Manual Processing of Unstructured Content Supported by ML






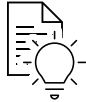


# USE CASE 2: Extracting Insight from Unstructured Text Using LLMs



Once the data has been pre-processed into a machine-readable format, the team then evaluates, tests, and deploys models.



## GenAI / LLM's

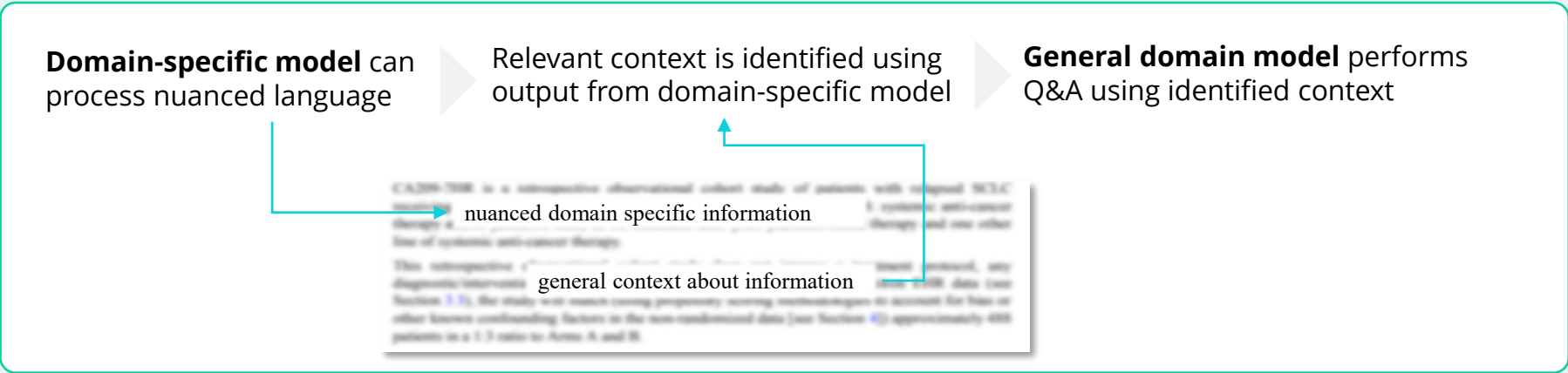
- 
**EXPLORE VARIOUS LLMs**  
 to identify best model for business problem  FLAN-T5
- 
**IDENTIFY FIT-FOR-PURPOSE MODELS**
  1. Selected a pre-trained **general domain model** to support Q&A tasks
  2. Selected a **domain-specific, name entity recognition model** that is pre-trained on large relevant datasets
- 
**EXTRACT INSIGHTS**  
 from unstructured text, leveraging the identified models
- 
**HUMAN SME REVIEW**  
 Insights are reviewed by human SMEs for verification
- 
**INFORMATION IS MORE ACCESSIBLE TO SUPPORT ANALYSIS**  
 Making information more available for analysis ultimately can improve CDER's ability to meet mission goals.

### LOOKING FORWARD

**FINE TUNE MODEL:** The team is looking to start training a model on domain data



# USE CASE 2: Extracting Insight from Unstructured Text Using LLMs



# CDEROne Analytics: **Moving Forward**

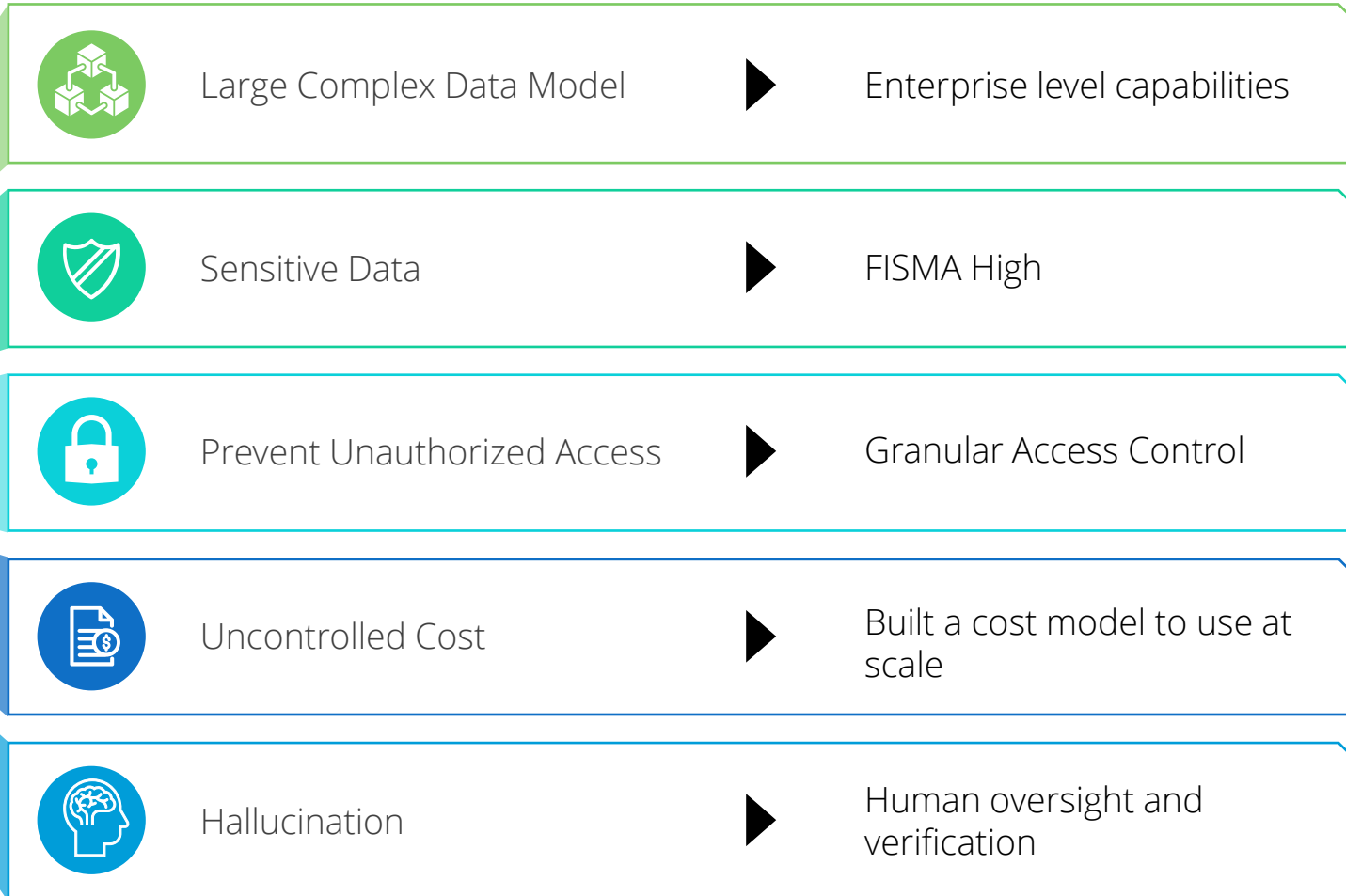


# Addressing Challenges in CDEROne



## CHALLENGES

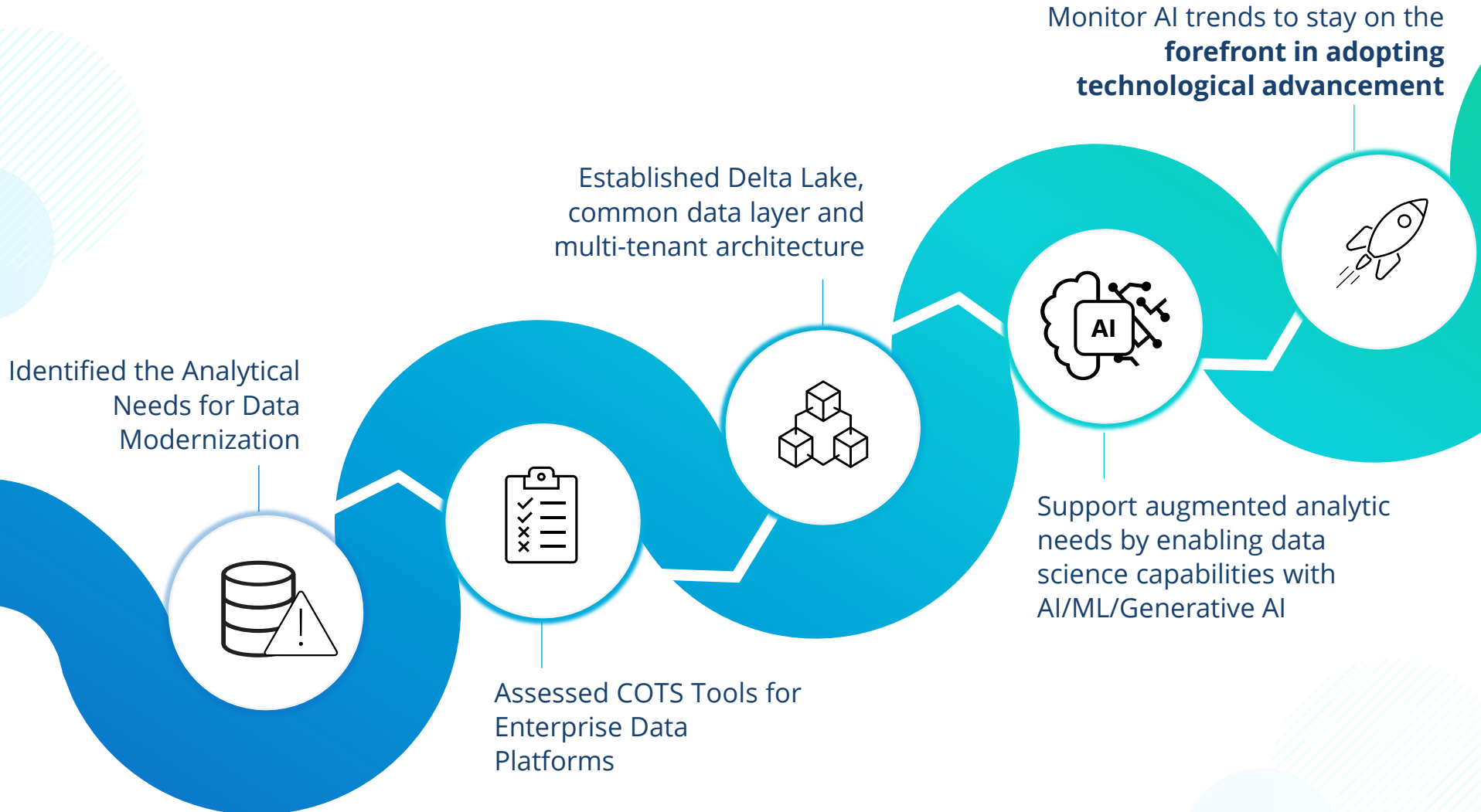
## SOLUTIONS



FDA and HHS is  
**sponsoring**  
**Databricks**  
**FedRAMP and**  
**FISMA High!**

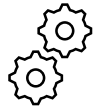
# CDEROne's Journey

How CDER leverages Databricks to support its augmented analytics needs throughout its regulatory process



# Supporting CDER's Mission With AI

Many opportunities are available for CDER to leverage AI advancement to support its mission goal from premarket regulatory review, compliance, pharmacovigilance, quality surveillance and prevent drug supply chain disruption.



**Enhancing FDA data** with AI for analyses that facilitate **regulatory assessments**

*GDUFA III Priorities*



Exploring use of AI to support completeness assessments of **prospective ANDA applications**



Leverage AI to **identify submissions containing RWD/E** and applications leveraging innovative clinical trial design

*PDUFA VII Priorities*



Build **models to identify supply chain disruptions** and potential shortages

*Drug Shortage Challenges*

“ Eventually our reviewers, compliance assessors and inspectorate, and the industries they regulate, **will use AI in everything from data collection to regulatory assessments**, coding and post-market surveillance. For us to be most effective and do our jobs as protectors of public health, it is essential that we **embrace these groundbreaking technologies**...to improve the chance that they will be **applied effectively, consistently, and fairly**.

**Dr. Robert Califf**, FDA Commissioner





**Thank you**