How CDEROne is Advancing CDER's AI Innovation Mission

CDER Office of Business Informatics
December 2023

Agenda

1. What Is CDEROne?
2. CDEROne's Data and AI Capabilities
3. Achievements in AI at CDEROne
4. Moving Forward

What is CDEROne?

What is CDEROne?

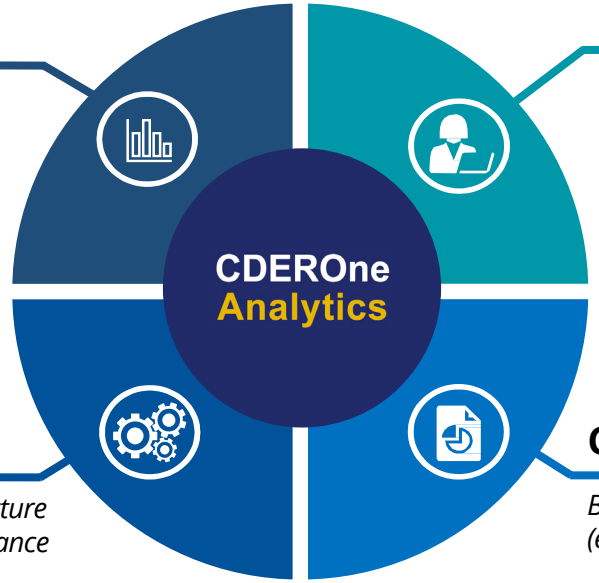
CDEROne Analytics is a secure Enterprise Platform that acts as a centralized one-stop-shop, integrating several data sources to provide tailored, self-service, augmented analytical solutions that support the business needs of the Center for Drug Evaluation and Research (CDER).

Advanced Analytics

Offers a conglomeration of advanced analytical tools & technology

Multi-Tenancy

Enables collaboration across all CDER offices via multi-Tenant workspaces



Enterprised Architecture

Provides a modernized architecture on a secure FISMA high Governance Cloud hosted in AWS

Centralized Data

Brings and connects CDER's data (external and internal) together

CDEROne Users

EXECUTIVES · ML ENGINEERS · DATA SCIENTISTS · DATA ANALYSTS · REVIEWERS

CDEROne Analytics

QUICK FACTS

3700+ Users Onboarded
to CDEROne Analytics

2500+ Hours saved
for Drug Safety Officers

45+ Use Cases
across CDER offices

12+ Tenants
across CDER offices

Why CDEROne is the Right Platform to Build CDER's AI Capabilities

- To support CDER's drug regulatory review process, data across a variety of domains needs to be accessible at an enterprise level.

CDEROne possesses foundational attributes for enabling AI capabilities:

- Data Abundance
- High Compute Platform that is Scalable and Accessible to Data Scientists
- AI capabilities (In-House Code Development + Open-Source Integration)
- Secure (FISMA high, Zero Trust)
- Platform support for citizen data scientists (i.e., help desk)
- Governance structure to prioritize development

There is a growing need for AI capabilities as CDEROne continues to scale up in its utility

CDEROne's Data and AI Capabilities

CDEROne's Data Layer

From ingestion to the publish layer, centralized data engineering serves as the foundation to build AI capabilities.

DATA SOURCES

Industry Submissions

FDA internal data generated through the review process

FDA published data (e.g., Orange Book, Purple Book)

External data (e.g., drug sales)

COMMON DATA LAYER

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

Data Ingestion → Data Processing, Management & Governance → Data Publication & Advanced Analytics

Supported by: Data Storage Tools, OCR Tools, Integrated Development Environments, Data Engineering Tools, Data Visualization Tools

TENANT SPECIFIC DATA ENGINEERING

This infrastructure allows for self-service data pipelines at a tenant/use case level

CDEROne's AI Layer

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

AI LAYER **Advanced Analytics tools and technologies made available to CDER data scientists.**

Flexible Programming Language → Model Building Frameworks → Centralized Collaborative Platform → AI Assisted Review Solution & User Interfaces

Supported by: Integrated Development Environments, Machine Learning Training, Registry, and Tracking Tools, Software Development Project Management Tools, Data Visualization Tools

Use case specific data on an as-needed basis

Insights & model outputs

COMMON DATA LAYER

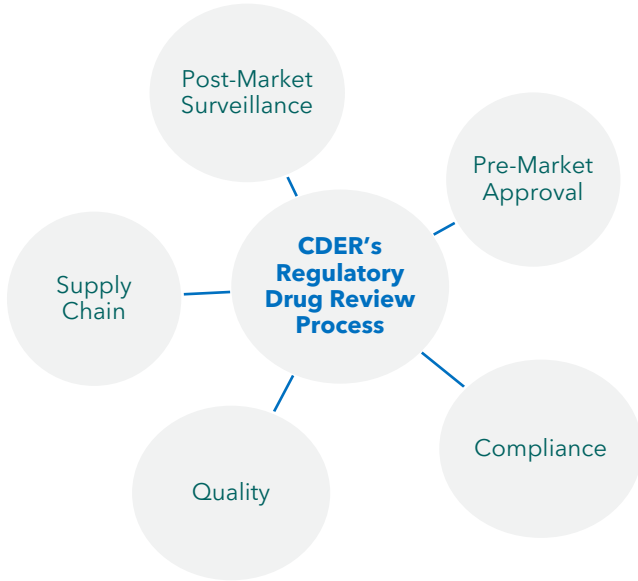
End-to-end data pipeline with centralized data engineering and analytics tools and technologies

DATA SOURCES

Data Ingestion → Data Processing, Management & Governance → Data Publication & Advanced Analytics

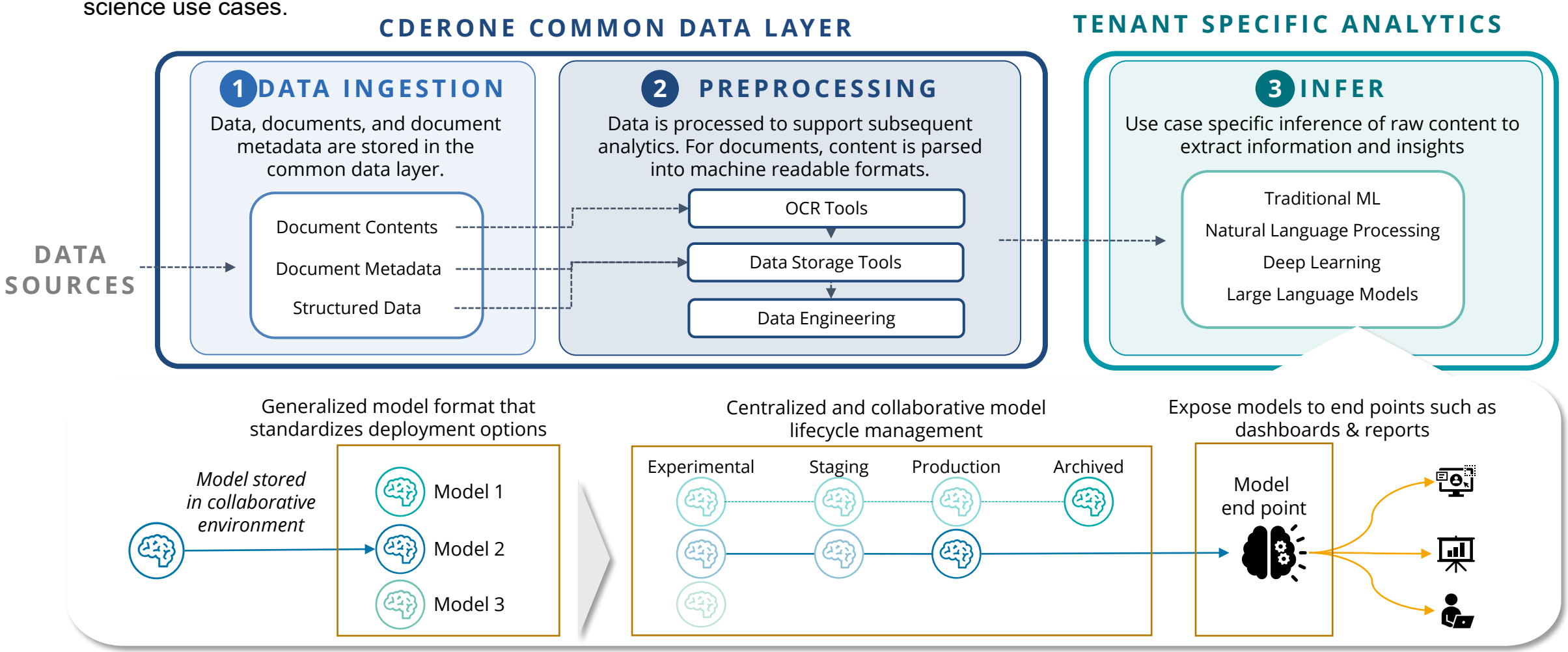
Business Process Improvements

This platform supports a variety of business processes by developing insights and supporting work across CDER Offices



CDEROne Architecture

CDEROne tenants can leverage a common framework composed of reusable capabilities to accelerate the document centric data science use cases.



What is AI and existing CDEROne use cases

Artificial Intelligence (AI) is an interdisciplinary area demonstrated by machines operating human tasks such as speech recognition, computer vision, translation between languages, etc.

Artificial Intelligence (AI)

The theory and development of computer systems able to perform tasks normally requiring human intelligence.

Machine Learning (ML)

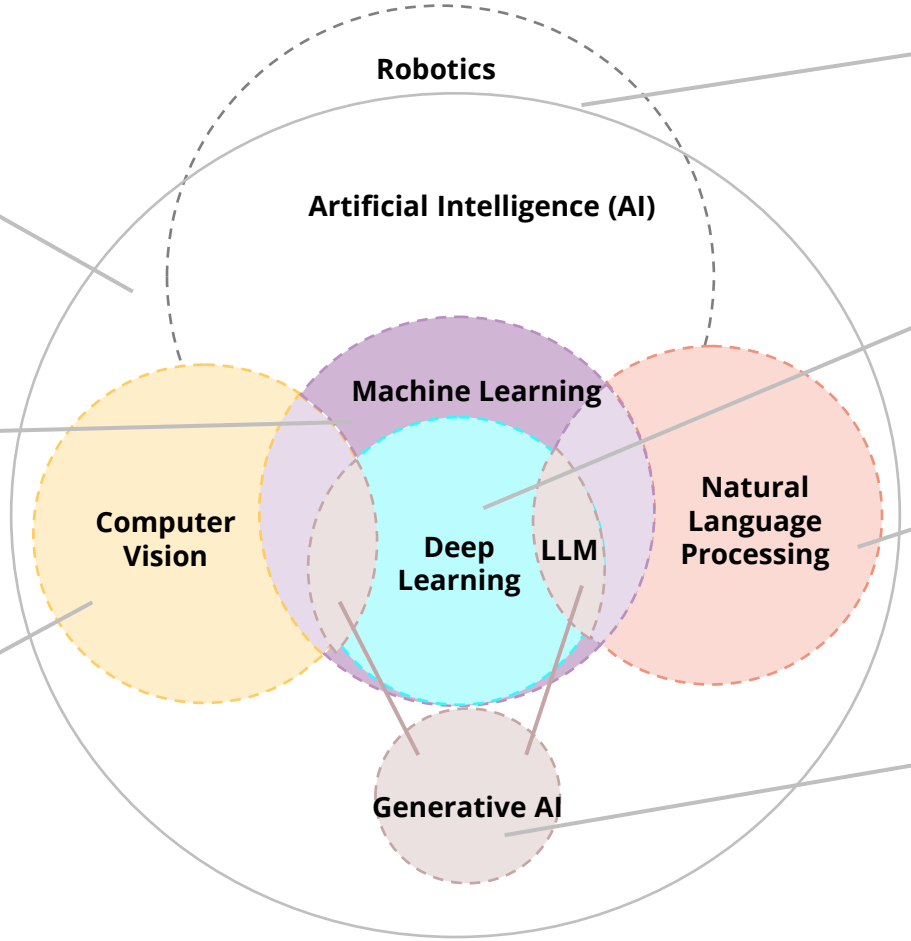
A field devoted to building methods that let machines "learn" leveraging data.

- **CDER Capability:** ML Classification & Operationalization

Computer Vision

A field that focuses on enabling computers to identify objects and people in images and video.

- **CDER Capability:** Text extraction using Optical Character Recognition (OCR)



Robotics

An interdisciplinary branch designing machines that can help and assist humans.

- **CDER Capability:** Intelligent Document Processing using Robotic Process Automation (RPA)

Deep Learning (DL)

Machine learning algorithms with brain-like logical structure of algorithms called artificial neural networks.

Natural Language Processing (NLP)

An interdisciplinary subfield of linguistics, computer science, and AI such as how to program computers to process and analyze large amounts of natural language data.

Generative AI (including LLM)

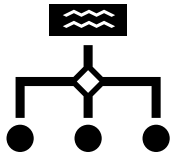
A type of AI system capable of generating text, images, or other media using generative models such as large language models (LLMs).

- **CDER Capability:** Extract insights from unstructured documents

Achievements in AI at CDERONe

How AI Can Improve CDER's Business Processes

AI can improve business processes by improving the accuracy, quality and accessibility of data, modernizing business processes, and creating new insights, among other improvements.



STRUCTURE UNSTRUCTURED DATA

FDA receives submissions and data in a variety of formats. Structuring this data and storing it in a central location makes it easier to leverage.

e.g., Computer Vision, Machine Learning, NLP

IMPROVE DATA QUALITY

AI can be used to validate, clean, detect anomalies, and ensure accuracy and consistency across large data sets.

e.g., Machine Learning, NLP, Deep Learning, LLMs

STREAMLINE OPERATIONS

These models are aiding manual processes to improve accuracy and efficiency.

e.g., Machine Learning, Robotic Process Automation

GENERATING NEW INSIGHTS

AI can create insights that is difficult to achieve by manual processes or traditional analytics methods.

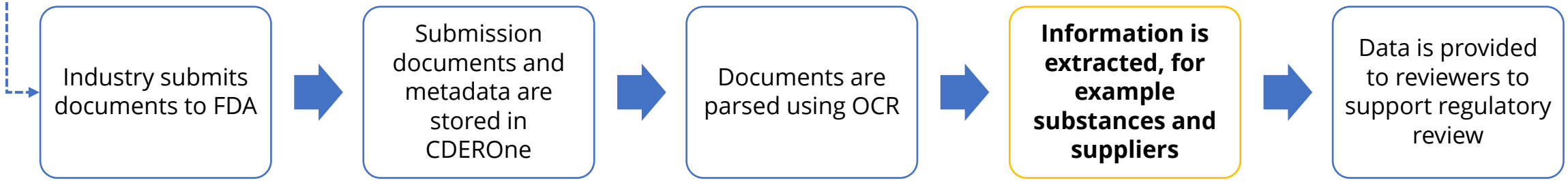
e.g., NLP, Deep Learning, Generative AI (including LLMs)

Generating Insights from Unstructured Data Using AI

Valuable data is provided to FDA in unstructured formats, making it difficult to access and use.
AI can be leveraged to structure and extract information, making it usable for reviewers.

Regulatory Review Process

Business Problem: Node B plays a critical role in the review process. In this node, critical information on substances is buried within unstructured documents. FDA employed advanced analytics to identify and contextualize these substances.



At this step, CDEROne's AI layer uses OCR, Deep Learning, and Large Language Models to structure data and gleam insights.

Added Capability

Generating Insights from Unstructured Data Using AI (Cont'd)

DATA SOURCES

Unstructured data in document

COMMON DATA LAYER

Document and document metadata is stored in common data layer

Document content is extracted into a machine-readable format

AI LAYER

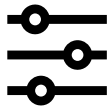
Use case specific inference of raw content to extract information and insights

Now that the data has been pre-processed into a machine-readable format, it can be passed into models. The team then evaluates, tests, and deploys models to identify and contextualize substances.

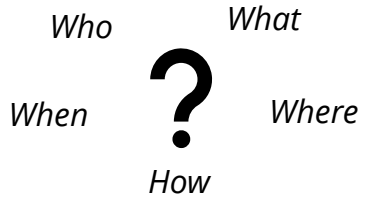
Explore Models: Test a variety of models to identify the best model for the business problem

Select model: Selected a name entity recognition model that was pre-trained on large data sets to identify substance names

Fine tune model: Retrain model on CDER data and evaluate for accuracy



Contextualize Data: Use a large language model to understand context about the substance

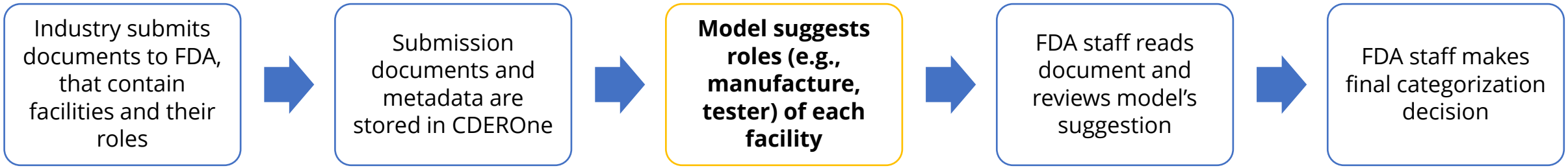


Impact: This effort structured unstructured data, filling in data gaps and improving data quality, ultimately generating new insights, which improved CDER's ability to:

- Conduct surveillance
- Understanding a fuller picture of layers in the drug manufacturing process

Supporting Manual Processes Using Machine Learning

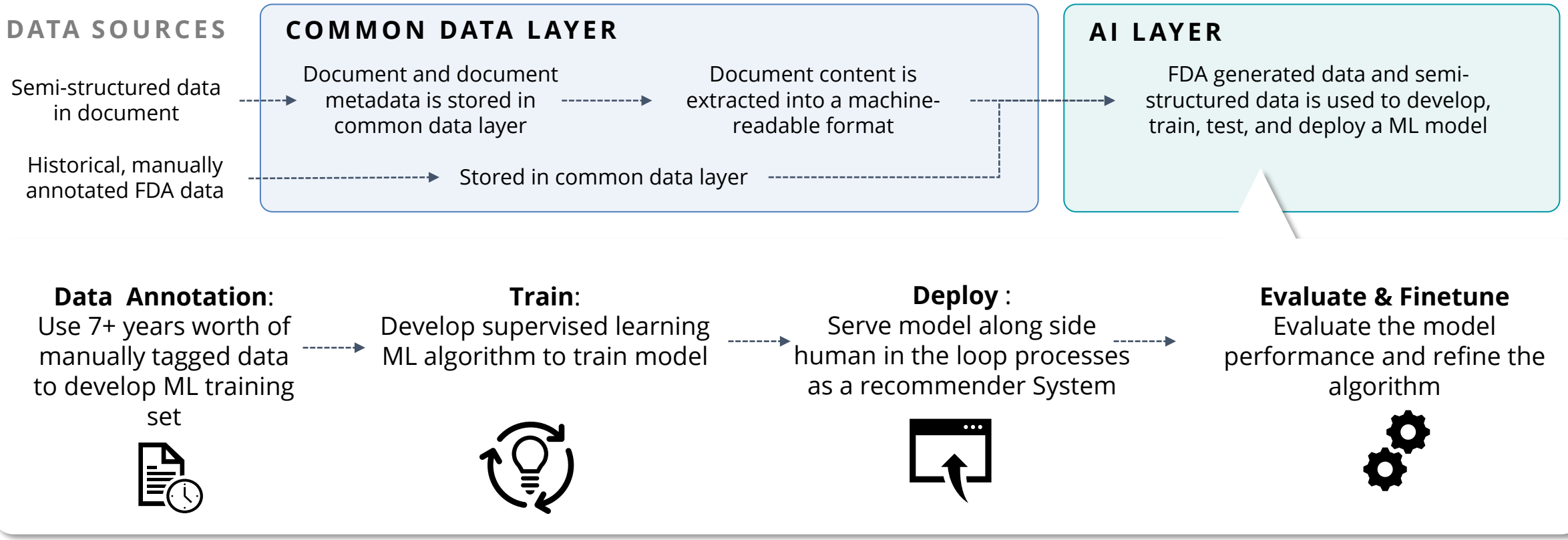
Leveraged AI to support the manual process of classifying industry submitted information (e.g., facility business operation).



At this stage, machine learning is used to help optimize this manual process.

Added Capability

Supporting Manual Processes Using Machine Learning (Cont'd)



Impact: The manual process of reviewing and categorizing information is now aided with machine learning, improving accuracy and efficiency.

Data Scientists Leverage AI Capabilities in CDEROne

Created a self-service environment to carry out data science projects to solve **policy and operational** challenges

Policy or Operational Challenges

Client has
policy question
or
operational challenge

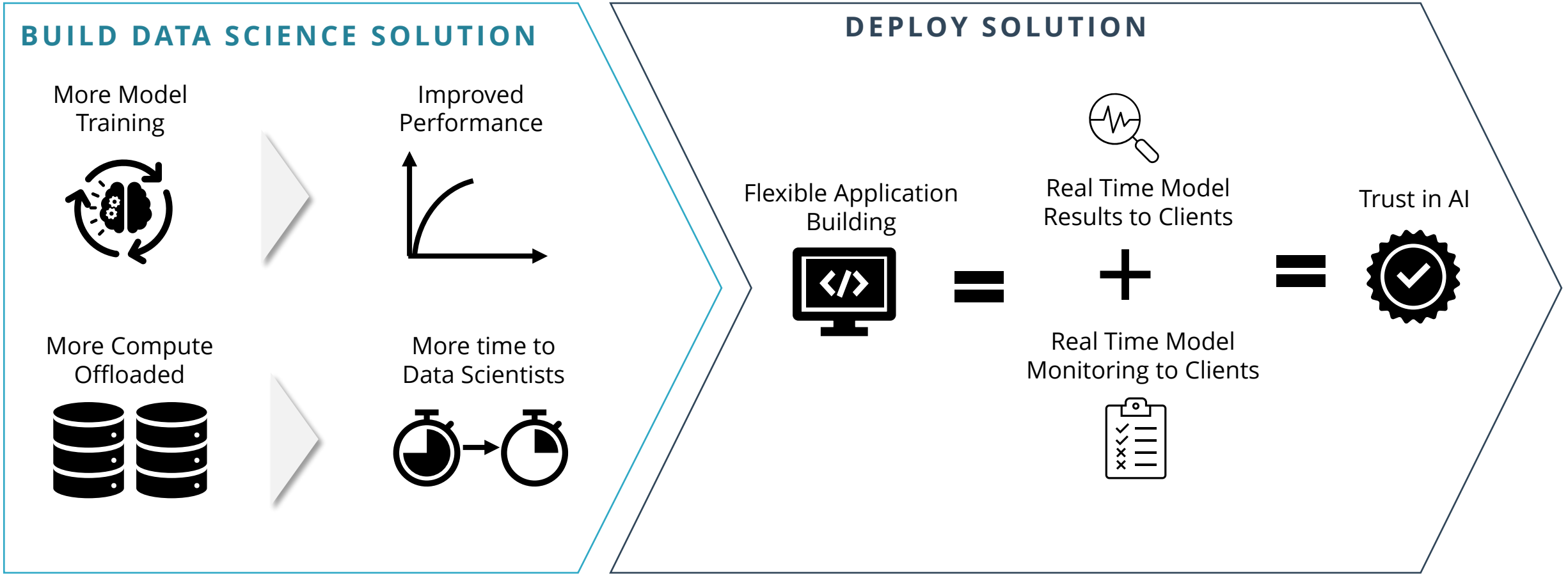
Build Data Science Solution

CDEROne provides
self-service data science environment to build AI models

Deploy Solution

CDEROne provides
self-service deployment
to host AI powered applications

Data Scientists Leverage AI Capabilities in CDEROne (Cont'd)



Impact: 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 clients **increasing trust in AI models.**

Moving Forward

Moving Forward Outline

Infrastructure

Enterprise level capabilities require scalable, high-performance infrastructure that can handle large datasets and complex models.

Security

Adhering to FISMA high Governance Cloud standards and implementing a Zero Trust is paramount to protecting sensitive data.

Access Control

Defining granular permissions to prevent unauthorized model and data access and continuing to implement a governance policy down to the data attribute level.

Computational Cost

Current pricing models could lead to uncontrolled costs for organizations when used at scale.

Cross Cloud Platforms

Integration, data interoperability, and consistent performance across cloud environments requires careful orchestration.

Keeping these known challenges in mind, CDEROne plans to expand its state-of-the-art analytics through..

Enterprise Level AI

Expanding to be a ready-to-use enterprise level AI capability that can support citizen data scientists. Continue to identify the data needs and the AI capabilities needed to be included as a part of the enterprise capability.

Enhance Open Search Capability

Leverage advanced artificial intelligence and natural language processing techniques to enhance multi-dimensional contextual/semantical searches.

Strengthen the AI Community of Practice

Collaborate with AI CoP in providing the necessary platform to identify, prioritize, and help develop AI related use cases, in support of CDER's mission.

Thank you!