2024 GENERATIVE AI JOURNEY

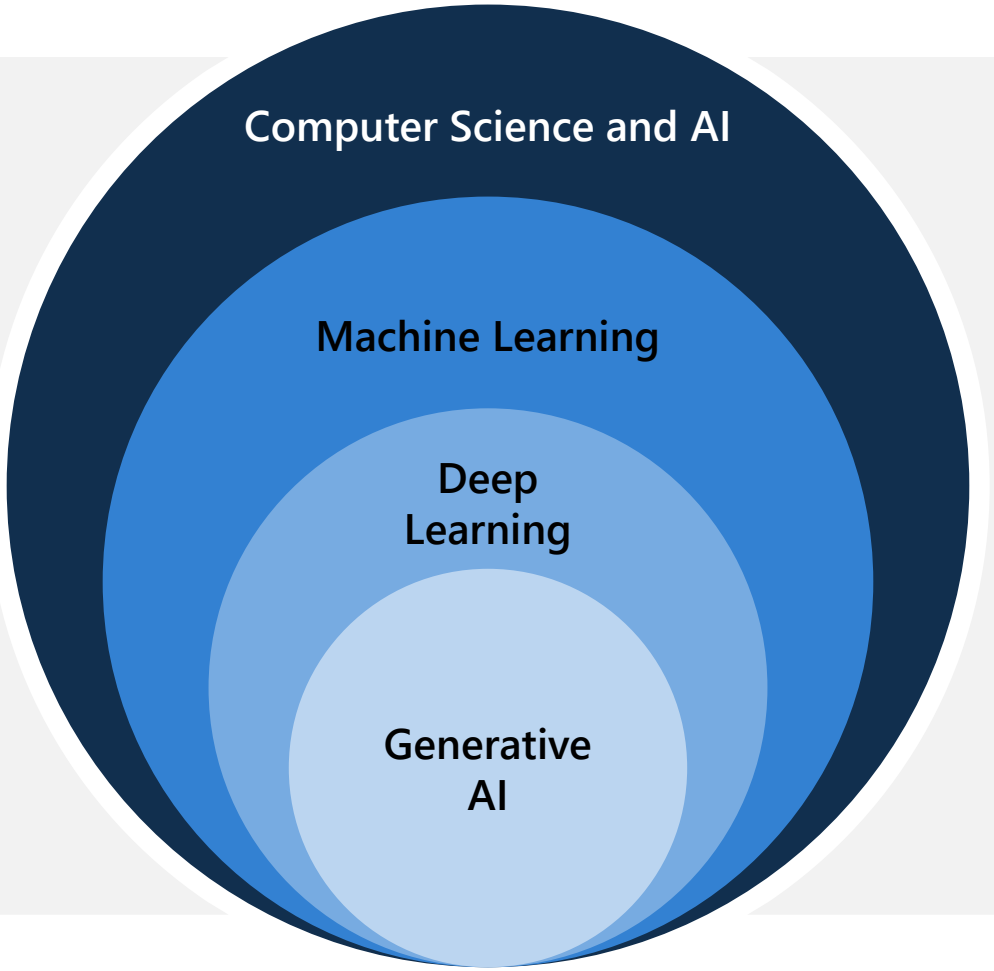
Joseph Raetano

AI Architect

VIP Continued Education Series 2025

01/07/2025

The Generative AI Journey



1956

Computer Science and Artificial Intelligence

AI is a field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence.

1997

Machine Learning

Subset of Computer Science / AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions.

2012

Deep Learning

A machine learning technique in which layers of neural networks are used to process data and make **decisions.**

2021

Generative AI

Create new written, visual, and auditory content given prompts or existing data. Transformer models gave way to LLMs.

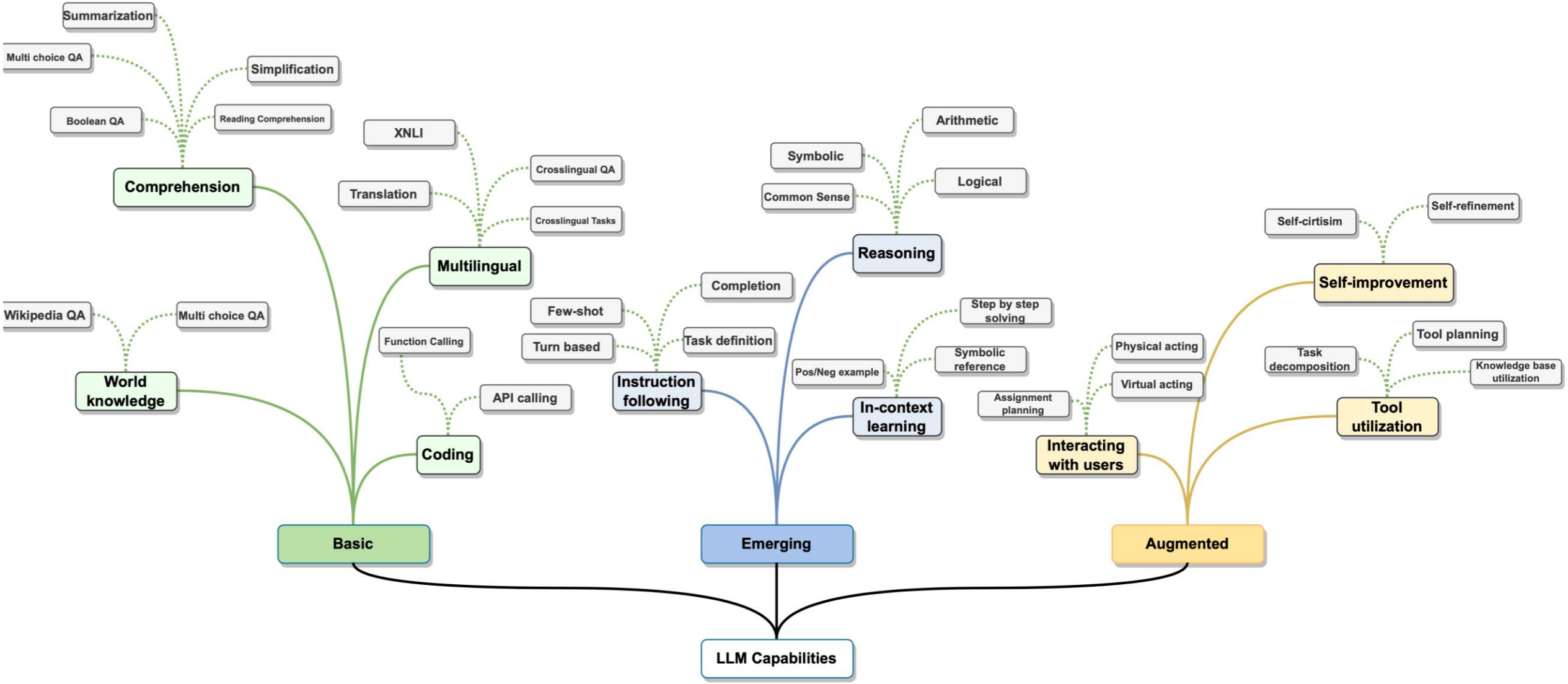


Fig. 1: LLM Capabilities.

The Generative AI Journey at VA 2024

- Azure OpenAI Service
- Azure AI Services
- Azure AI Search
- AI Language Service
- Text Analytics For Health
- Azure Machine Learning
- Azure Databricks
- CosmoDB

Feb

Azure OpenAI and related AI Services GA in Feb

Azure OpenAI, AI Search, AI Services, etc. were made generally available in February 2024 on Microsoft Azure Government (MAG).
Initiated the SDP AI/ML Community of Practice (CoP)

Jun

These Services went FedRAMP High in June

With these services at FedRAMP High enabled them quickly on SDP and got moving on production AI use-cases

Sep

Enabled VA's first GenAI enabled data pipeline

Enabled CLEVER Pipeline with Azure OpenAI

Oct

Information Assistant (IA) enabled in Dev/Test

Enabled IA, a Microsoft Chatbot, RAG, Agent template

Dec

Completed IA Security Impact Assessment

Completed IA SIA can now move to production in 2025
Completed 36 AI/ML CoP Sessions

Where Are You in AI Maturity?

MATURITY LEVELS		1	2	3	4	5
		Initial/Ad Hoc Individual Project	**Repeatable Team Project**	**Defined Program**	**Managed Portfolio**	**Optimized Enterprise**
OPERATIONAL AREAS	**AI Ops**	• Reactionary ad hoc AI capability identification • Start of use-case analysis • Purchased proprietary non shareable AI services	• Initial process to capture AI product requirements and user workflows • Vendor supported and training identified	• Formal AI Product Management team, strategy and roadmap exists	• AI Product Management Goals are linked to organizational performance objectives.	• AI capability dependencies are mapped across organizational boundaries • New AI use-cases being developed, deployed and shared
	ML Ops	• Algorithms, models and methods are selected ad hoc and not documented	• Implemented standardized methods for documenting experiments • Standardized algorithm, ML model and methods selection	• Algorithm, model and methods catalog exists • Model to use case matching leverages historical knowledge	• Automated selection, testing, and evaluating ML models • AutoML achieved and unlocked	• Feedback from ML tools captured and models are CI/CD • AutoML enhanced speed • Framework maximized for continuous integration, improvement and deployment
	Data Ops	• "Shoeboxes" of data; stored locally, not discoverable, and copied from one machine to another	• Routine non standardized data sources available • Data discovery is ad hoc • Data catalog and pipeline needs identified	• Engineering support for data management activities is explicit • Data catalog and pipelines created • Data governance needs identified	• Self-service for adding new data sources and preparing datasets • Curating data for ML projects	• Intelligent, secure data discovery • Access of data across organizations • Metrics on business usage and compliance • Continuous Integration & Deployment
	Sec Ops	• Code security/validation is manually accomplished in the Test environment • Container security is default to orchestrations	• Code security/validation is manually accomplished within the pipeline • Container security is baselined at orchestration	• Code/Container security, validation is automated within the pipeline & manually approved • Established secure process for containerizing tools and moving into production environment	• Code/Container Security validation automatically approved; software rollouts are "trusted" • Validated secure automated process for deploying secure software	• Pipeline security software feeds central Security Data Lake; Automation embedded at the Pipeline Orchestration layers; Automated code rollbacks;
	Dev Ops	• Development on local workstation • Simple server instantiation for projects	• Process to moving locally developed tools into production • Some portions are still manual	• Established secure process for containerizing tools and moving into production environment • Utilizing dev, test, prod environment	• Increasingly automated process for deploying secure software with emphasis on reducing iteration and delivery timelines	• Fully-managed secure software container orchestration; CI/CD/CATO
	Cloud Ops	• Minimal Cloud Resources or Individual User Account	• Needs surfaced and documented • Innovation Sandbox created in the cloud • Talent & Training identified	• Dev, Test, and Prod environments created & available • Manual resource allocation	• Self-service or templated cloud resource allocation	• Balanced automated resource sharing across the organization with robust cost/benefit/usage metrics • CI/CD
	People Ops	• AI Employee journey created • Languages documented • KSA's identified • Successful agency projects identified	• Needs surfaced and documented • Talent map created • Training identified	• Enhanced recruiting strategies • Internal talent lifecycle mapping • Internal talent development • Internal/External communications development	• Utilize agile methodology • Utilize enhanced software tools that combine most HR related activities to include training	• AI sustainment cycle created • Innovation mindset promulgated

THE VA PRODUCT MANAGER'S JOURNEY TO AI-ENABLED CAPABILITIES

This map represents a common set of moments that product managers experience while trying to oversee the development of various AI-enabled capabilities. While this map does not represent what happens to every single product manager, it does identify moments when many experience pain points throughout the process. Those moments provide a guide for where VA Summit Team can focus its time and resources towards the goal of maximizing the experience with VA Summit Data Analytic Platform and helping to create more impactful AI enabled

Product Managers

JOURNEY

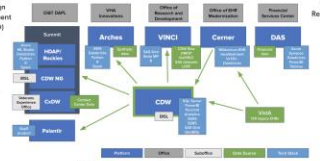
EXPERIENCE

BROKER IDEAS/USE CASES	APPLY FOR FUNDING AND WAIT	GET ACCESS TO ANALYTICS PLATFORM AND ACQUIRE DATA	AI PRODUCT DEVELOPMENT	TRANSITION
TIME: 1 MONTH	TIME: 3 MONTHS	TIME: 3-6 MONTHS	TIME: 3-6 MONTHS	TIME: 3 MONTHS

PRODUCT MANAGERS BEGIN HERE

FIND AND EVALUATE IDEAS · DS/AI Reference Catalog Idea Backlog · "SELL" IDEA TO BUSINESS UNIT · Create Memorandum of Agreement (MoA) · DESIGN AND HAVE WORKSHOP · DESIGN AND CREATE PILOT · MAKE DECISION TO MOVE FORWARD · Process Design Document (PDD) to Elicit Requirements · Statement of Work (SOW) · ACQUISITION SUPPORT · Request for Quote (RFQ) · SI Onboarding · Data Agreement · Data Card · ACQUIRE DATA · Solution Discovery Document (SDD) · SPECIFY ANALYTIC PLATFORM NEEDS · Solution Design Document (SDD) · Fill out Forms · Analytic Platform Account Creation · GET ACCESS TO ANALYTIC PLATFORM · Create Test Plan · Product Requirements · Project Status Reports · DEVELOP MODELS AND DEPLOY ENDPOINT OR INTO SOFTWARE · Change Request Forms · Authority to Operation (ATO) · TRANSITION TOOL · Training Profiles Updates · PLACE INTO PROGRAM PRODUCTION

Task Order Contract Large BPAs
- **PTEMS (System Integrators with Developer Support)**
- **Tis(Program/Project Managers)**
- **Best practices from Product Lines**



Emotional Scale

Product Manager collects ideas, assesses, triages, and prioritizes opportunities from the teamed Clinician (CN) and Data Scientist (DS)

- Information comes in ad hoc
- No centralized location to store ideas that come in so they can be tracked over time and can be analyzed for redundancy with other ideas
- Difficult to identify duplicative and/ or tangentially related ideas quickly
- No clear decision making framework to qualify product ideas

Data Scientists attempt to determine use-cases and analytic platforms to best utilize for any given use-case

- Difficult to sell idea to an office without a product, vision, committed VA investment, projected costs to office, cost-benefit to office and infrastructure requirements to accept new capability
- Users are often hidden behind various separate large organizations, making it difficult to engage and determine what their needs are
- No easily accessible centralized location to track data science community, their contact information, interactions and their needs for future reference
- It is difficult to work with healthcare data because it is constantly evolving and there is a large learning curve. Need clinical knowledge

Clinician (CN) as the SME brings ideas and helps to fill gaps. Looks for assistance from COE.

- They have data and ideas but unsure of how to operationalize it.

PMs develop a business case to obtain VA investment and resources

- Inefficient and ineffective creation of PDD due to lack of awareness of resources available to help with its development
- Unsure how to solicit help from internal teams to help identifying data requirements and strategy
- Has someone done this before or is there work being performed already like this
- Unsure of what types of data sources already exist within CDW that may expedite data acquisition
- Inefficient to develop process as it is passed around and back and forth through email to get feedback and agreement, sometimes with people editing outdated versions, leading to rework
- Inability to track progress or changes due to lack of routing mechanism and actions sent through emails get lost in a slew of other emails causing delays in the process

Data Scientist finds a lot of difficulties

- Lack of tools around knowledge sharing - like medical concepts, data integration, want to reuse things - library to expose this, align definitions
- Knowledge management has been a large problem, - "Shared millennium"- share libraries, code, documents. These are being shared on teams, SQL databases.
- Need an effective way for collaboration, one integrated datasource - patient data is not in a centralized location. Collibra not working out well..not intuitive
- Difficulty of working with healthcare data - constantly evolving data, need clinical knowledge, healthcare learning curve
- Access and resourcing of local machine/servers are limited

Once funding is approved by Acquisition (ACQ), PMs solicit contract support to develop AI-enabled capabilities and wait through a lengthy acquisition process

- Difficult to know when money will be available
- PMs feel there isn't a clear process to get the acquisition process started
- The internal acquisition process takes too long or is too rigid for smaller purchases
- Lack a good way to track overall funding and burn rates once funding is approved.
- It's difficult to determine the status of contract support and funding across multiple contracts
- Lack a central location to determine status, priority, funding decision and burn rate of contracts, must rely on searching through emails or outdated information that has changed.

Data Scientist (DS) works with the Data Engineers (DE) to determine data needs, sourcing, understanding restrictions, and any formal data agreements

- Difficult to track the process/progress of being onboarded onto platform or platforms - what's next, who has the action?
- Complete onboarding to avoid program delays
- When things go wrong with onboarding it is unclear who to contact
- Finding and assessing quality of data is difficult. Often a struggle to get operational data from sources other than CDW
- Unsure what type of data sources already exist other than on CDW
- There is a pretty clear process for getting access to CDW data & it just requires supervisor approval and then it is done
- Data in real environment or across services may look and be different , making it more difficult to clean and utilize, Synthetic data is often needed
- Not sure how to get data that is pulled from various places with their own individual formatting requirements structured in way that can easily be used
- If using shared data, it is difficult to determine who owns the data or if there are restrictions
- Difficult to determine what data is available in VA without access as users are unable to identify them prior to onboarding
- Difficulty to differentiate between various authoritative data sources.
- Who is our single point of communication when help is needed from a data engineer

PMs try to figure out how to start the development process within VA. Once they understand the process, they coordinate with a Data Analyst (DA) to describe and communicate environment needs to the analytic platforms team then wait for their environment to be stood up

- Unsure of where to get access to analytic platform, understanding and training for various platforms increases mental fatigue on team
- To many systems to analyze use-cases appropriately
- Unable to review requirements easily in a collaborative way
- Potentially redundant information from PDD is required for VA use case requirement documentation
- Use case documentation is passed around through email, making it difficult to keep track of the newest versions of information
- Difficult to analyze what tools are available for use for any given use-case within VA
- Difficult to determine appropriate approved substitutes for tools needed
- Time-consuming to get access to data analytic platform environments
- Unsure how to request changes to environment after original submission
- Need more people interested in data science and need to expand curiosity
- There needs to be coordinated training for each data analytics platform: entry level, intermediate level, advanced level

PMs fill out more documentation to ensure proper access to and tracking of analytic platform, Data Scientist gets moving on building mathematical models

- Once onboarded to a project, if they need to be added to other projects, the process is unclear.
- No clear process to have users removed from projects.
- Information may get lost because process relies on getting an email or a conversation to get a person added to specific projects
- How to track how a project is moving, is there a way to dashboard projects for any given analytic platform or platforms

Product Teams work on developing models with ML Engineer and productizes using the the various VA data analytic environments

- Unsure who to contact if there are issues with the environment or changes need to be made
- Unsure of where to go to get questions answered
- Unsure when and how to get other parts of the organization involved such as test and assess for V&V
- Unsure when or what new features are available for use
- Difficult to provide demos to offices for products as they are being developed on any given platform.
- Little time to keep stakeholders in the loop through the product development cycle to understand when they will need to plan for this product

PT/PM transitions AI capability to the Product Owner (PO)

- Environments and licenses aren't always compatible with the users environments or software and will need an ATO to use.
- It's unclear how OIT and product owners will be involved in model maintenance.
- Various customer/partners have different data structures or software license issues that may make it difficult to make models transferable
- Difficult to help users become comfortable with new capabilities and models.
- No standard way to transition models from data analytic platform to agency specific purpose.
- No authorized MLOps Pipeline, we have to recreate the transition process each time.
- Lack of knowledge of how MLOps works & who would maintain MLOps infrastructure
- Lack of standards for production ML models (transparent metadata, model cards, etc.)

Designed & Illustrated by Joseph Rastano DAS

AI/ML Community of Practice

In 2025, we aim to build on achievements in 2024 by focusing on the development and operationalization of AI products that further support the Department of Veterans Affairs (VA). This will involve expanding our RAG (Retrieval-Augmented Generation) pipelines, refining MLOps strategies, and exploring the latest advancements in multi-agent systems, generative AI, and data security. We will prioritize delivering tangible AI solutions that address VA-specific challenges, from healthcare decision-making to enhanced data analytics and operational efficiencies.

To foster targeted skill-building and meaningful outcomes, we recommend several new CoP session themes for 2025:

1. **AI Product Strategy & Management** – Best practices for transforming AI prototypes into VA-aligned, scalable products.
2. **Advanced Multi-Agent Systems & Automation** – Designing and orchestrating AI agents for complex, coordinated tasks across VA domains.
3. **Responsible & Compliant AI** – Techniques and frameworks for bias reduction, robust security, and regulatory compliance in VA AI deployments.
4. **MLOps at Scale** – Building enterprise-grade pipelines that emphasize governance, reproducibility, and real-time performance monitoring.
5. **Vision & Advanced Modalities** – Integrating computer vision and emerging modalities, such as GPT-Vision, into RAG workflows and clinical support tools.

Work Only | Generative (Ungrounded)

Clear chat | Adjust | Info

Chat with your work data

Information Assistant uses AI. Check for mistakes. Transparency Note

Ask anything or try an example

Are there any radio telescopes in Australia?	What are Microsoft's primary sources of revenue?	What are some flavors of Breyers?

please write an abstract for

Adjust | Regenerate

Work Only | Generative (Ungrounded)

Clear chat | Adjust | Info

Chat directly with a LLM

Information Assistant uses AI. Check for mistakes. Transparency Note

please write an abstract for

Adjust | Regenerate

Information Assistant, built with Azure OpenAI

Chat Manage Content Math Assist (preview)

Work Only | Generative (Ungrounded)

Chat directly with a LLM

Information Assistant uses AI. Check for mistakes. Transparency Note

please write an abstract for

Adjust Regenerate

Information

Build and Version

1.2

Azure OpenAI

Instance

vac20sdpoai899c22197

GPT Deployment Name

gpt4o

GPT Model Name

gpt-4o

GPT Model Version

2024-05-13

Embeddings Deployment Name

ada

Embeddings Model Name

text-embedding-ada-002

Embeddings Model Version

2

Azure AI Search

Service Name

infoasst-search-agewv

Index Name

vector-index

System Configuration

System Language

English

Close

Upload Files | Upload Status

Supported file types

The Information Assistant copilot template currently supports the following file types:

Data
xml, json, csv, tsv, txt

Productivity Software
pptx, docx & xlsx

PDF
For page count maximum check documentation here

Web
htm & html

Email
eml & msg

Folder Selection

Create new folder

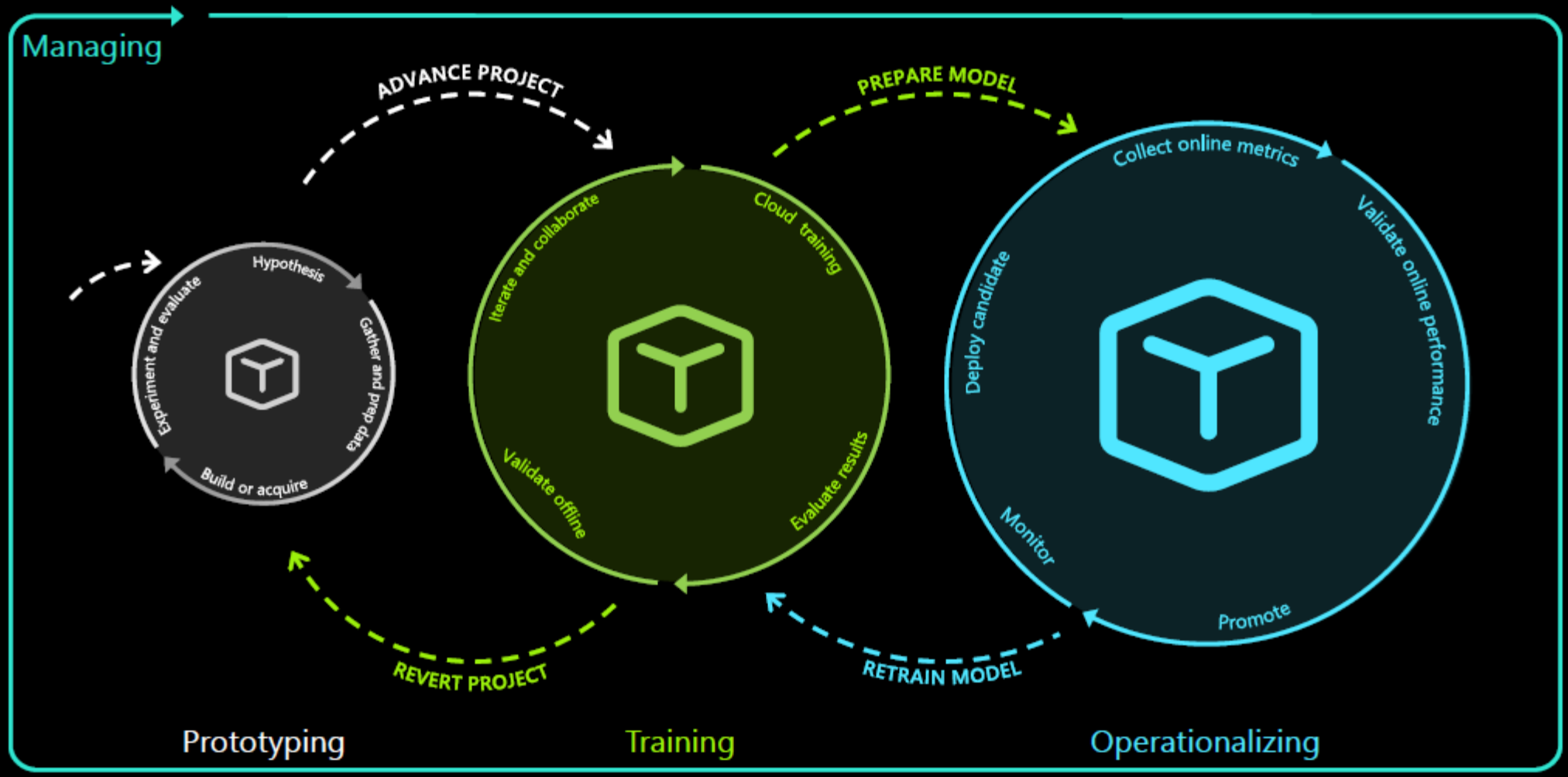
Tags

Click to Add files

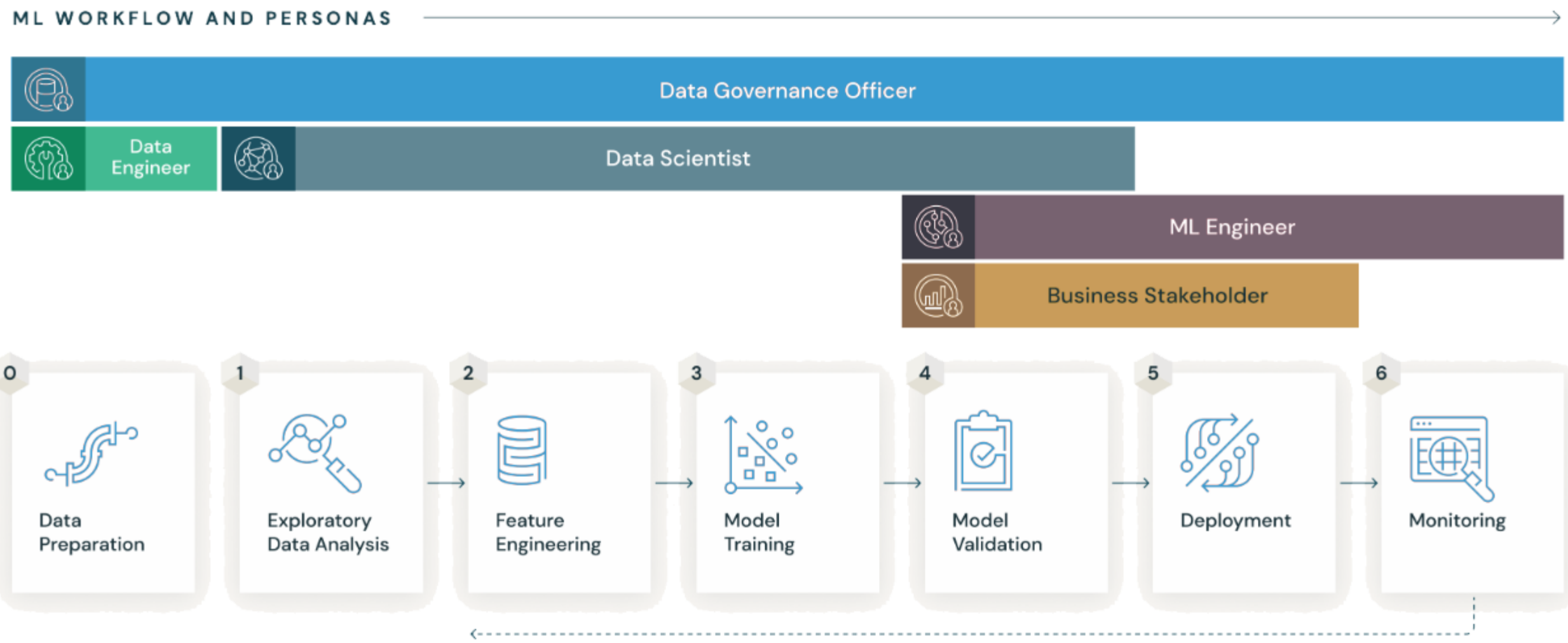
Or

Drag and Drop files here

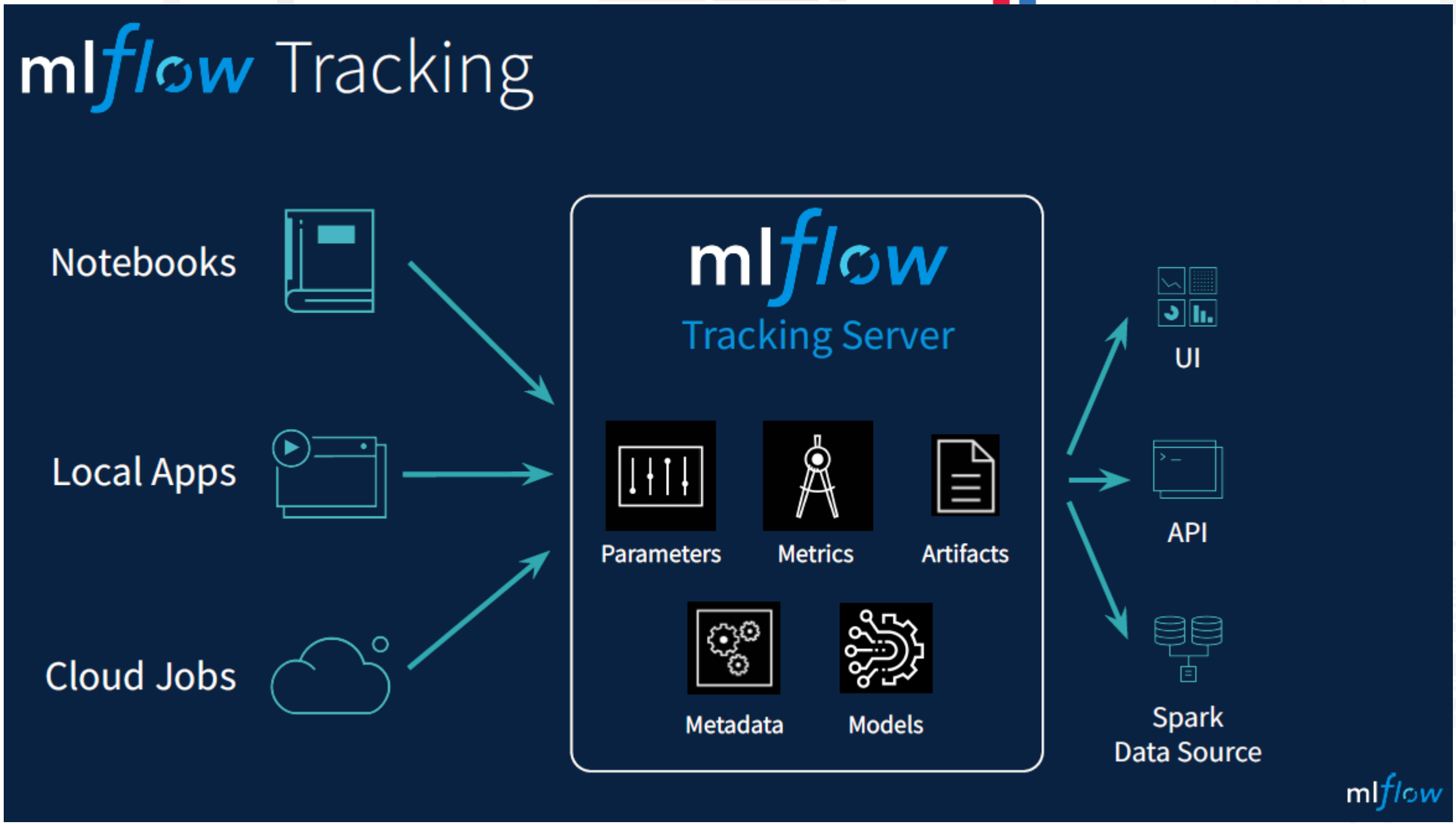
Enterprise Machine Learning Lifecycle in the Real World



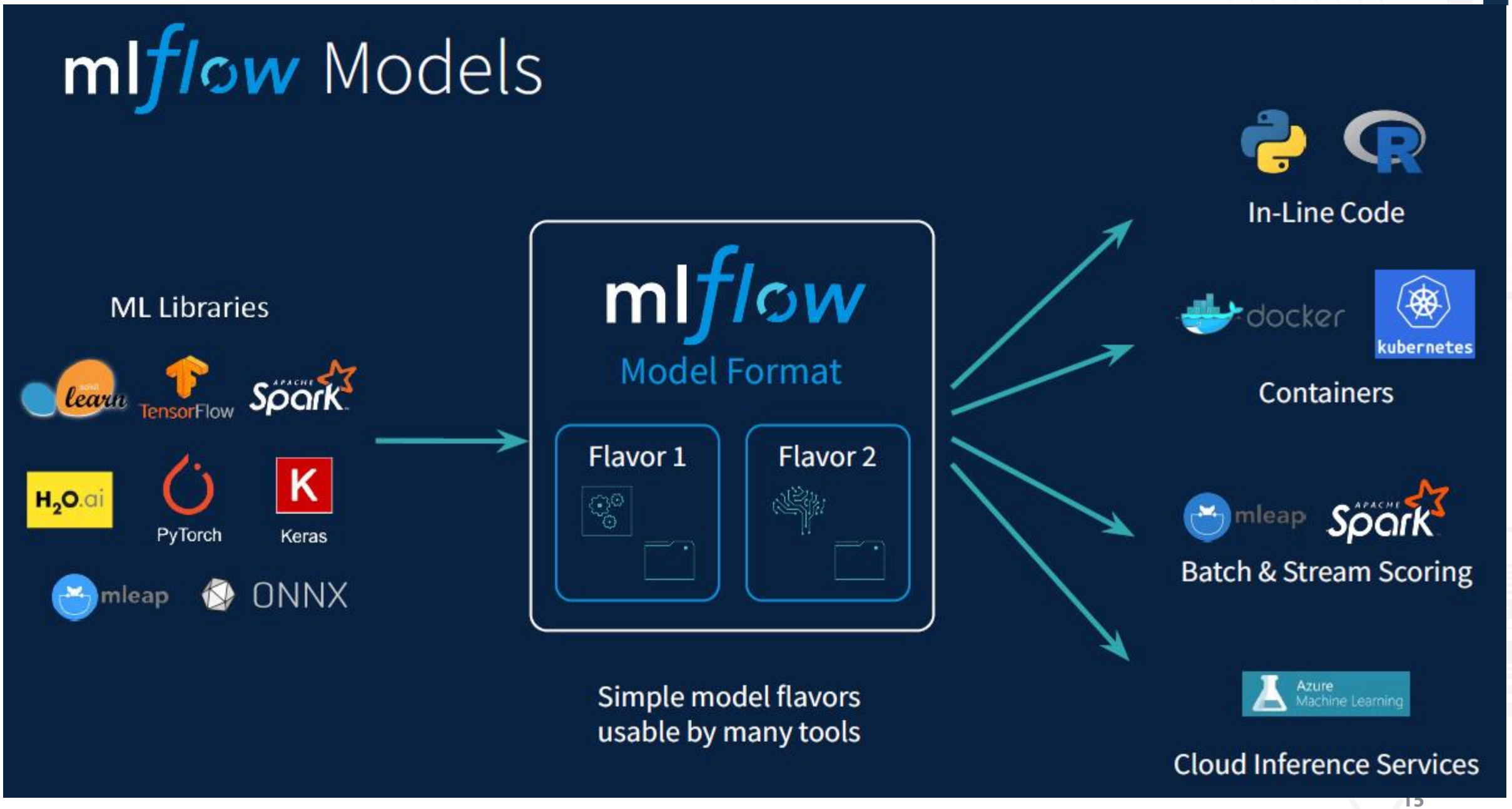
MLFlow

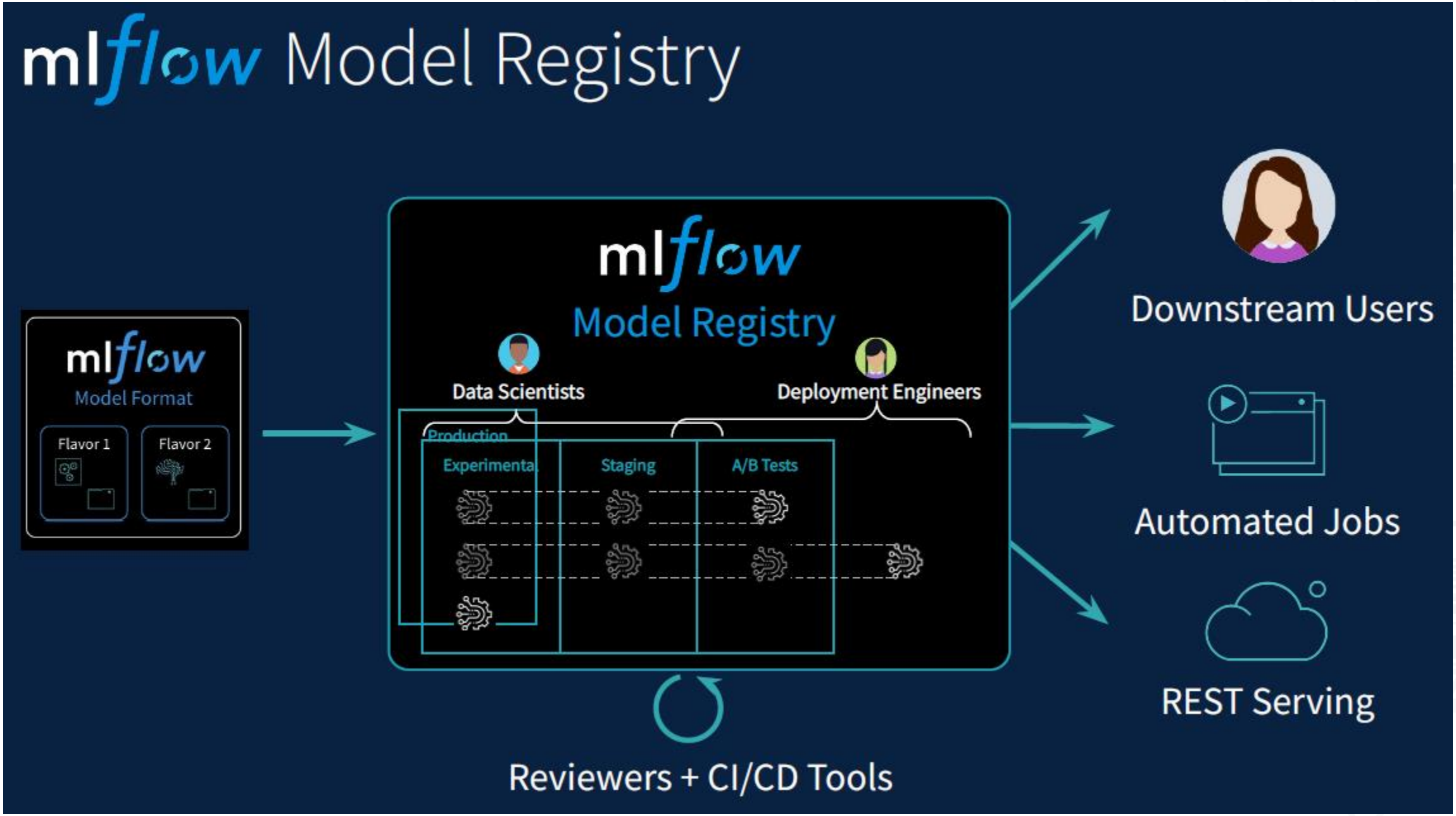

ML WORKFLOW AND PERSONAS
Data Governance Officer
Data Engineer
Data Scientist
ML Engineer
Business Stakeholder
0
Data Preparation
1
Exploratory Data Analysis
2
Feature Engineering
3
Model Training
4
Model Validation
5
Deployment
6
Monitoring

mlflow Tracking

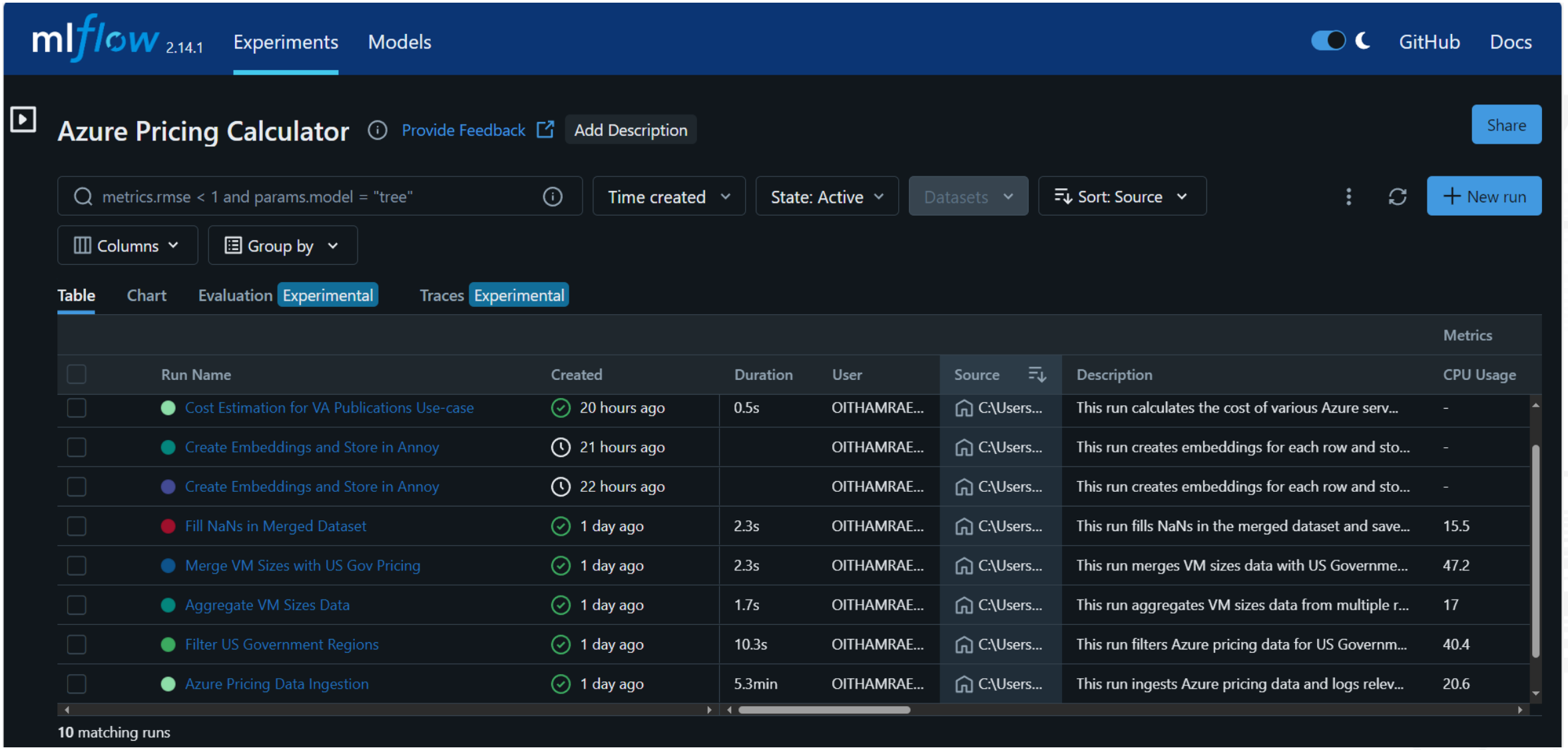


mlflow Models




mlflow Model Registry
mlflow
Model Format
Flavor 1
Flavor 2
mlflow
Model Registry
Data Scientists
Deployment Engineers
Production
Experimental
Staging
A/B Tests
Reviewers + CI/CD Tools
Downstream Users
Automated Jobs
REST Serving

Azure Pricing Calculator Experiment Runs



QUESTIONS?

THANK YOU