

# MLOps = DevOps<sub>4</sub>AI

Conf. dr. Cristian KEVORCHIAN

University of Bucharest

# Typical problems for ML projects

- <https://xkcd.com/1838/>



# Problems with the “pile of scripts” approach

- No traceability of work
- No clear artifacts created
- No governance
  - What data has been used?
  - What code?
  - What code + data resulted in my model?
  - What was the result for the model X?
- No reproducibility
- Not possible to hand over

Why my model  
does not give the  
same great result  
it gave me just  
yesterday?

Nothing have  
changed!

# Machine Learning Operations(MLOPS)

- ❑ MLOPS is based on DevOps principles and practices that increase the efficiency of workflows. For example, continuous integration, continuous delivery and deployment.
- ❑ MLOps applies these principles to the machine learning process, with the goal of:
  - Faster experimentation and development of models
  - Faster deployment of models into production
  - Quality assurance and end-to-end lineage tracking

# Team Data Science Process

- The Team Data Science Process (TDSP) is an agile, iterative data science methodology that enables the efficient delivery of predictive analytics solutions and intelligent applications.
- TDSP improves team collaboration and learning by recommending how team roles should be combined.
- The TDSP incorporates best practices and structures from Microsoft and other industry leaders to assist in the successful implementation of data science initiatives.
- **The goal is to assist businesses in fully realizing the benefits of their analytics program.**

# Key components of the TDSP

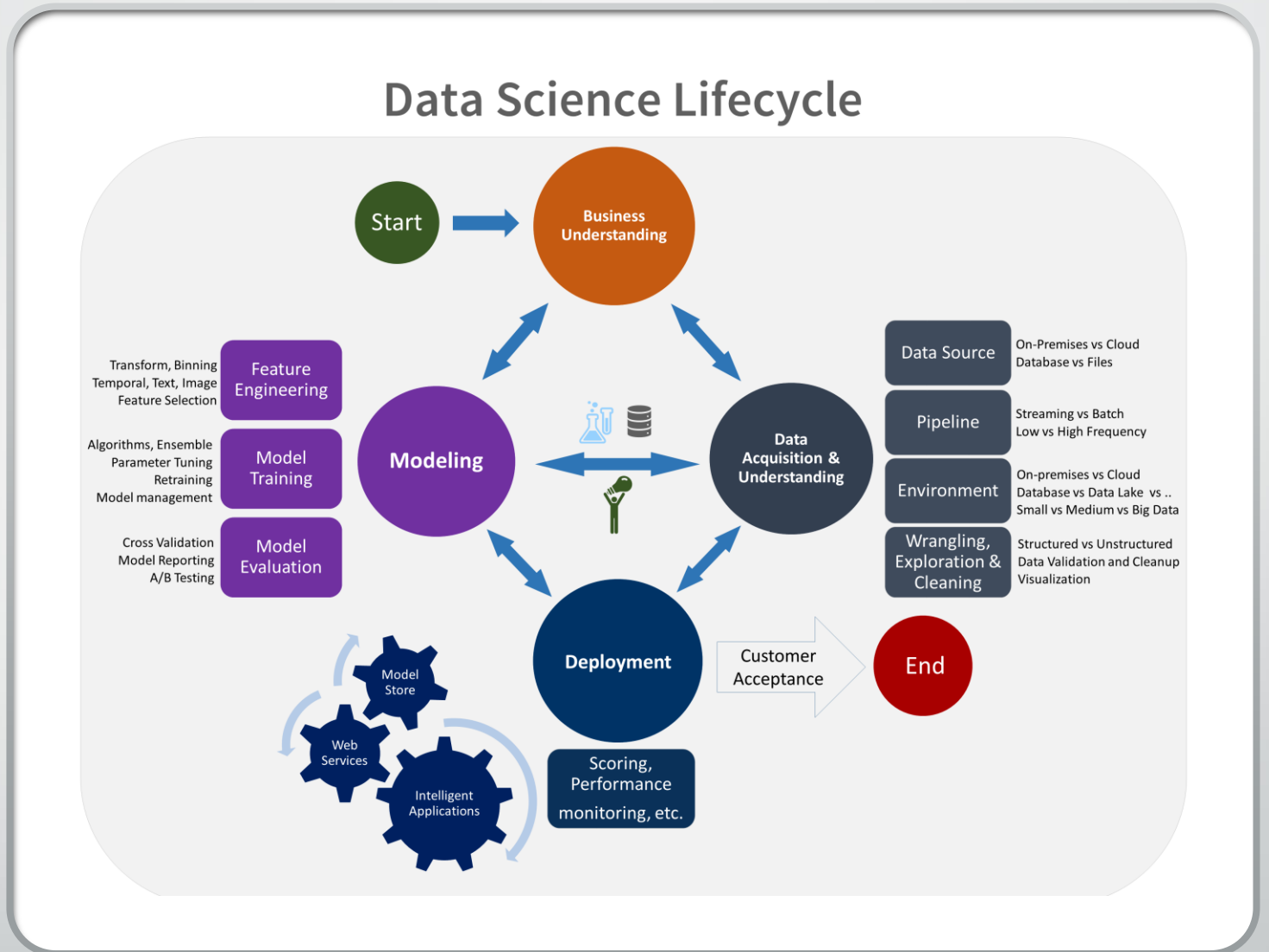
- Data science lifecycle definition
- A standardized project structure
- Infrastructure and resources recommended for data science projects
- Tools and utilities recommended for project execution

# Data science lifecycle

- The Team Data Science Process (TDSP) establishes a lifecycle for the development of your data science projects. The lifecycle outlines the entire process that successful projects must go through.
- This lifecycle is intended for data science projects that are delivered as part of intelligent applications. For predictive analytics, these applications use machine learning or artificial intelligence models. This process can also be used for exploratory data science projects or improvised analytics projects. However, in such cases, some of the steps outlined may not be required.

# ML Projects lifecycle

- **Business Understanding**
- **Data Acquisition and Understanding**
- **Modeling**
- **Deployment**







# ML Project Roles

- Solution architect
- Project manager
- Data engineer
- Data scientist
- Application developer
- Project lead

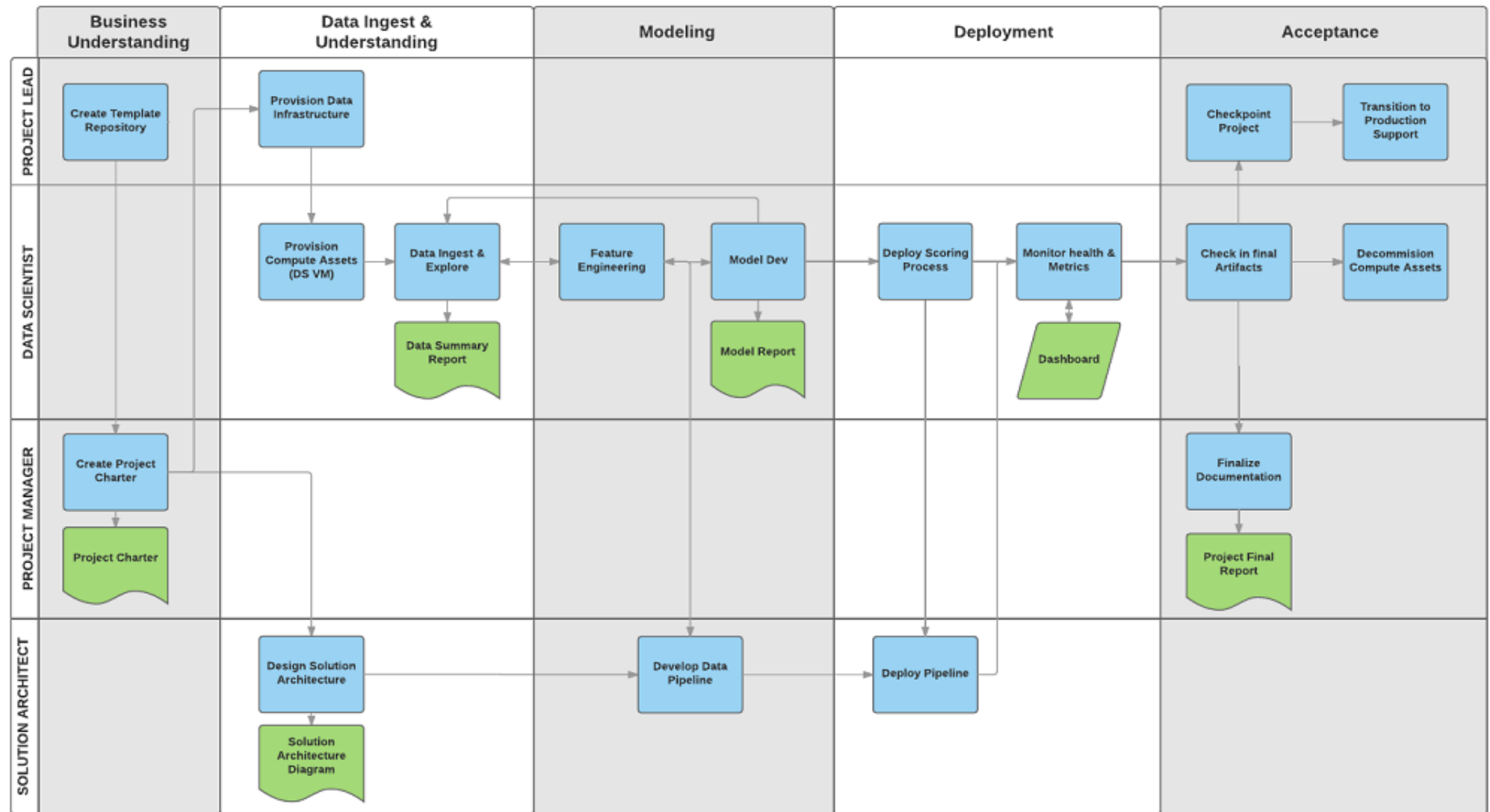
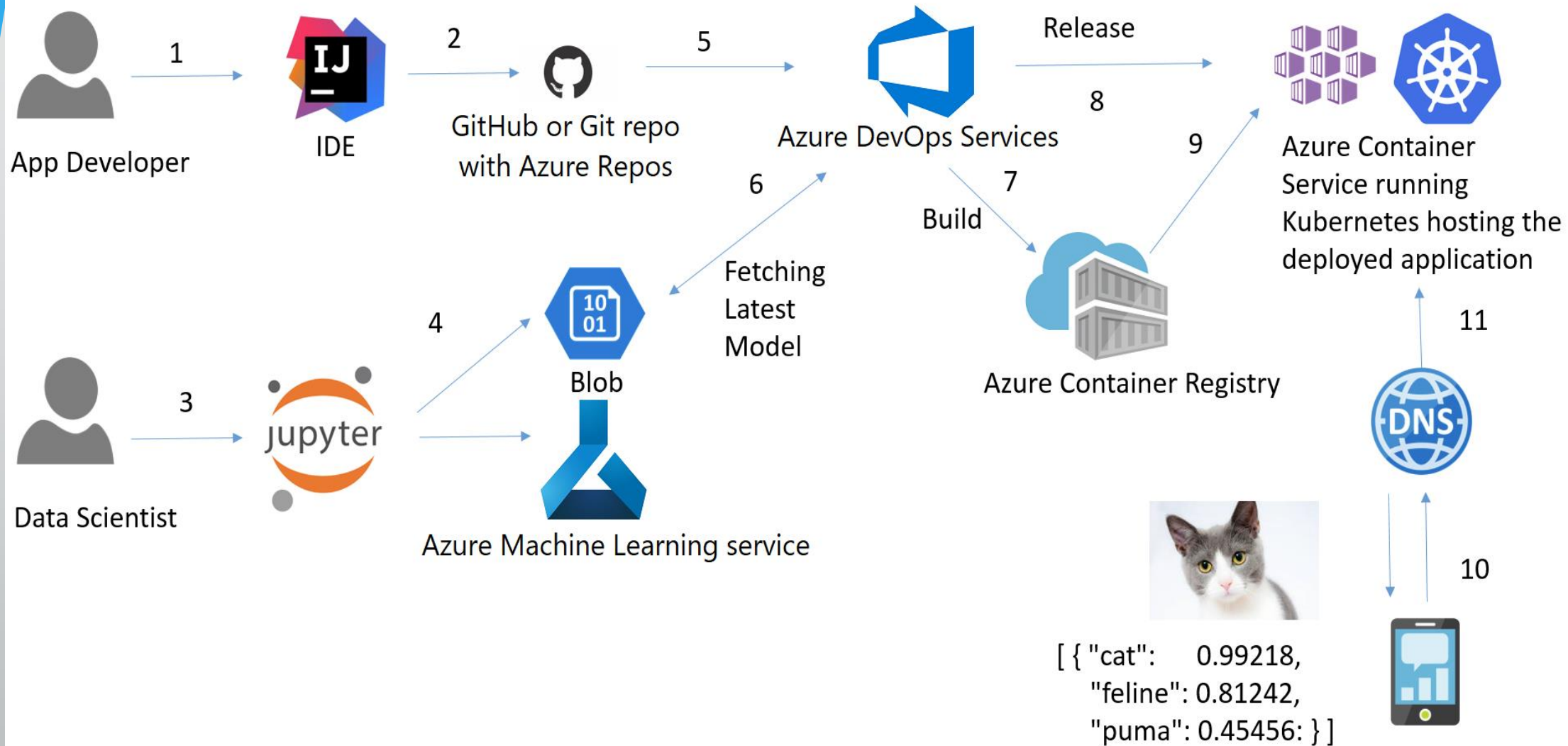


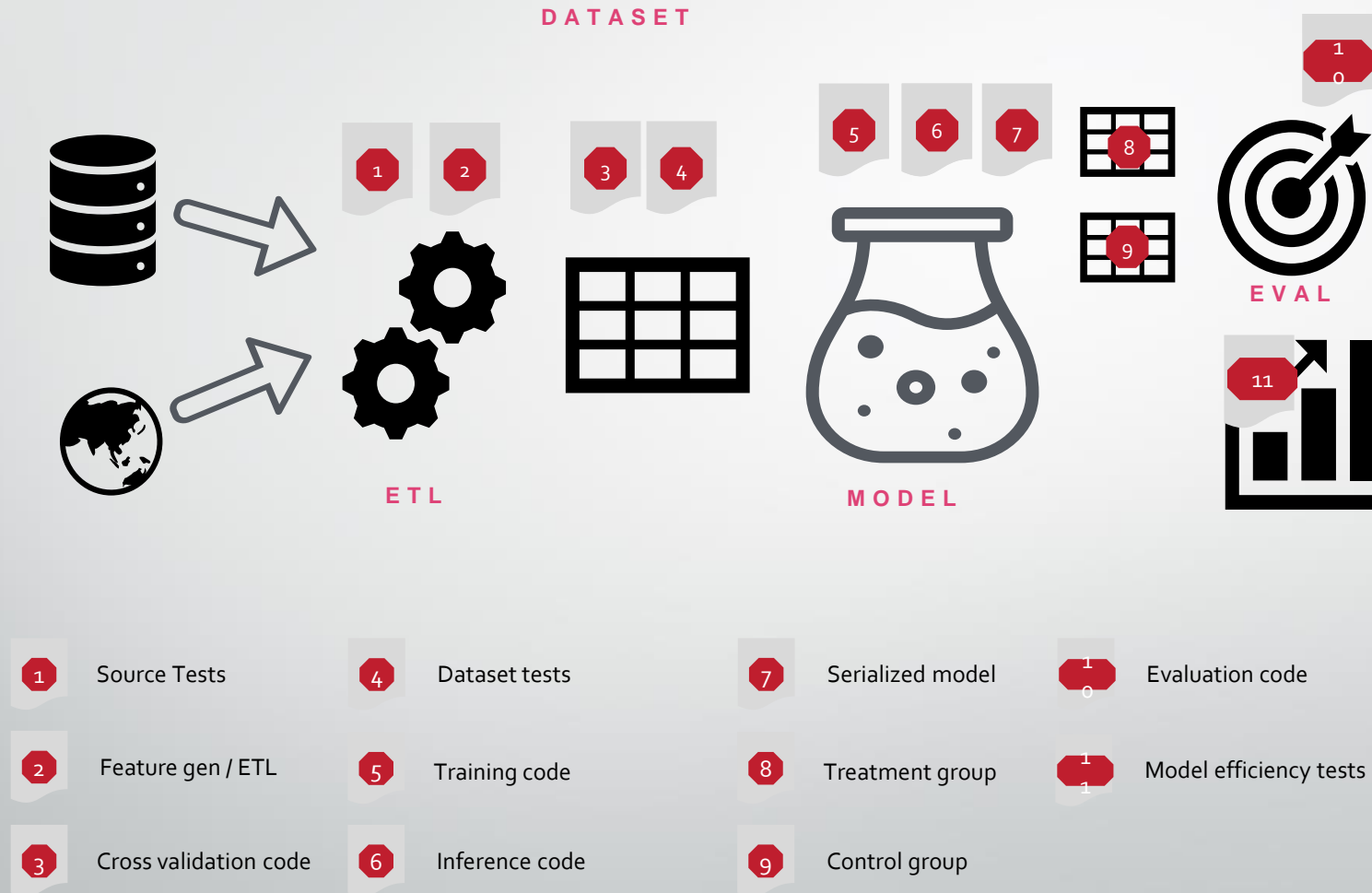
Diagram presents the tasks (in blue) and artifacts (in green) associated with each stage of the lifecycle (on the horizontal axis) for these roles (on the vertical axis).





# Developing Prod-ready ML

# What makes an ML model (asset examples)



# MLOps Benefits

## Reproducibility / Auditability

---

- Code drives **generation** and **deployments**
- Pipelines are **reproducible** and **verifiable**
- All artifacts can be **tagged** and **audited**

## Validation

---

- SWE best practices for quality control
- Offline comparisons of model **quality**
- Minimize **bias** and enable **explainability**

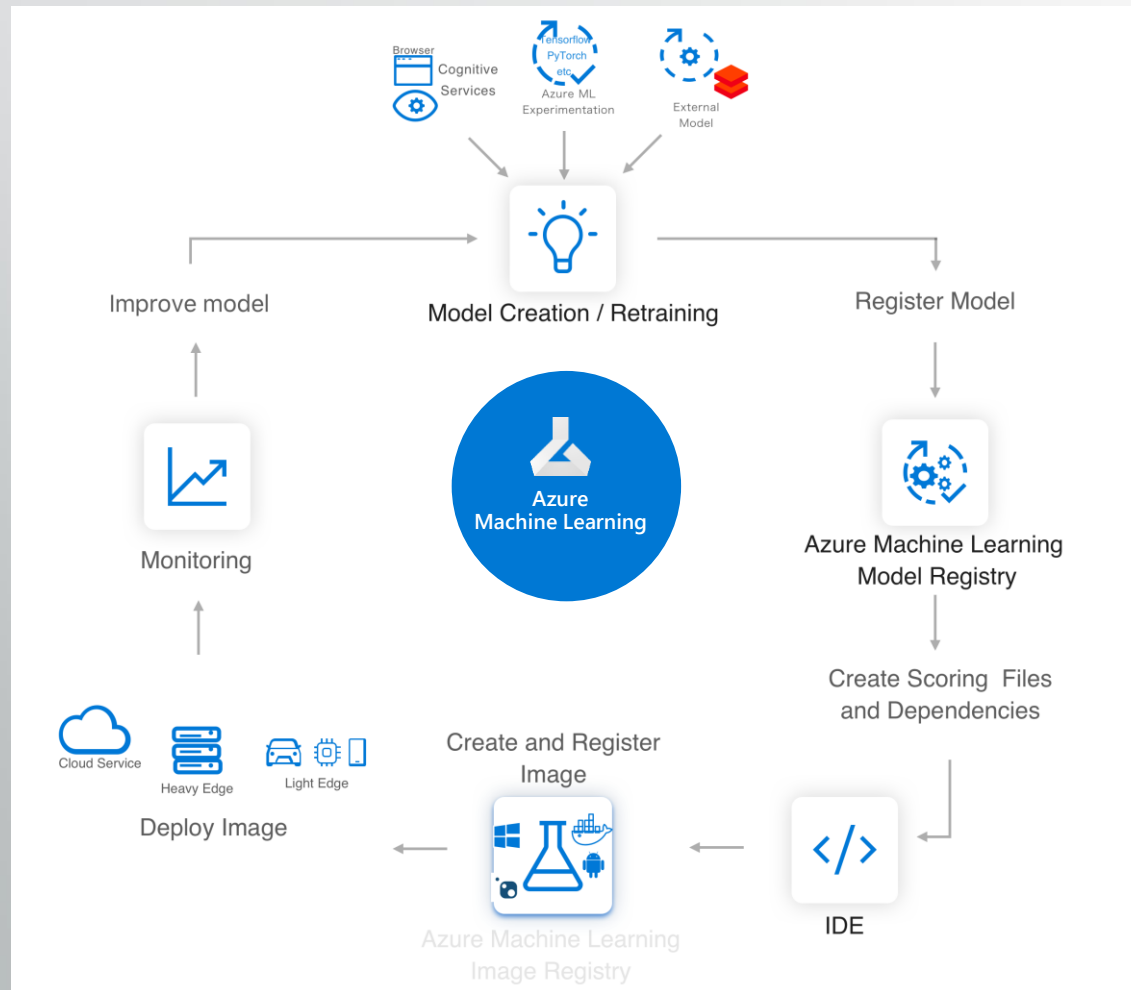
## Automation / Observability

---

- Controlled rollout capabilities
- Live comparison of predicted vs. expected performance
- Results fed back to watch for drift and improve model

# Azure ML service

Lets you easily implement this AI/ML Lifecycle



## Workflow Steps

Develop machine learning training scripts in Python.

Create and configure a compute target.

Submit the scripts to the configured compute target to run in that environment. During training, the compute target stores run records to a datastore. There the records are saved to an experiment.

Query the experiment for logged metrics from the current and past runs. If the metrics do not indicate a desired outcome, loop back to step 1 and iterate on your scripts.

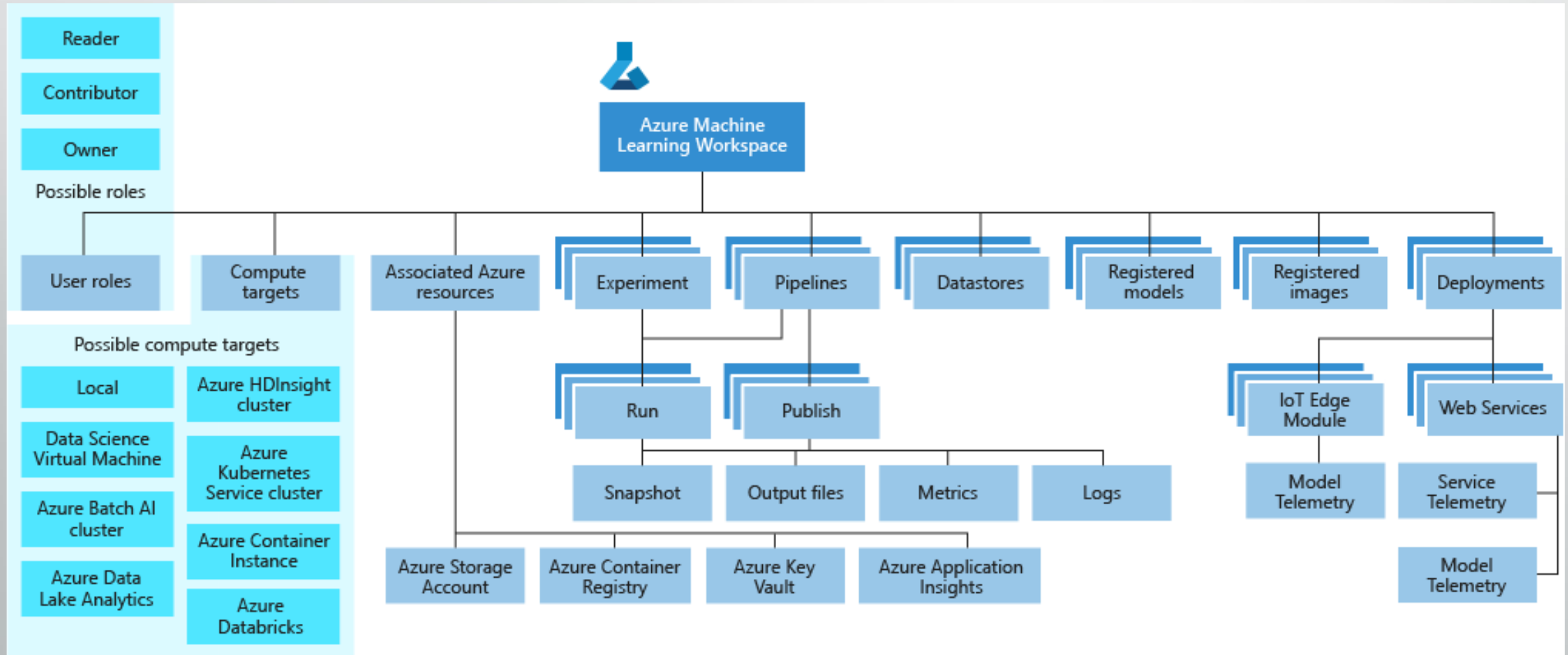
Once a satisfactory run is found, register the persisted model in the model registry.

Develop a scoring script.


Create an Image and register it in the image registry.

Deploy the image as a web service in Azure.

# Azure ML service Workspace Taxonomy







**“Best kept secret of AI is  
how it is de-abbreviated.  
Today we are going to  
reveal it.”**

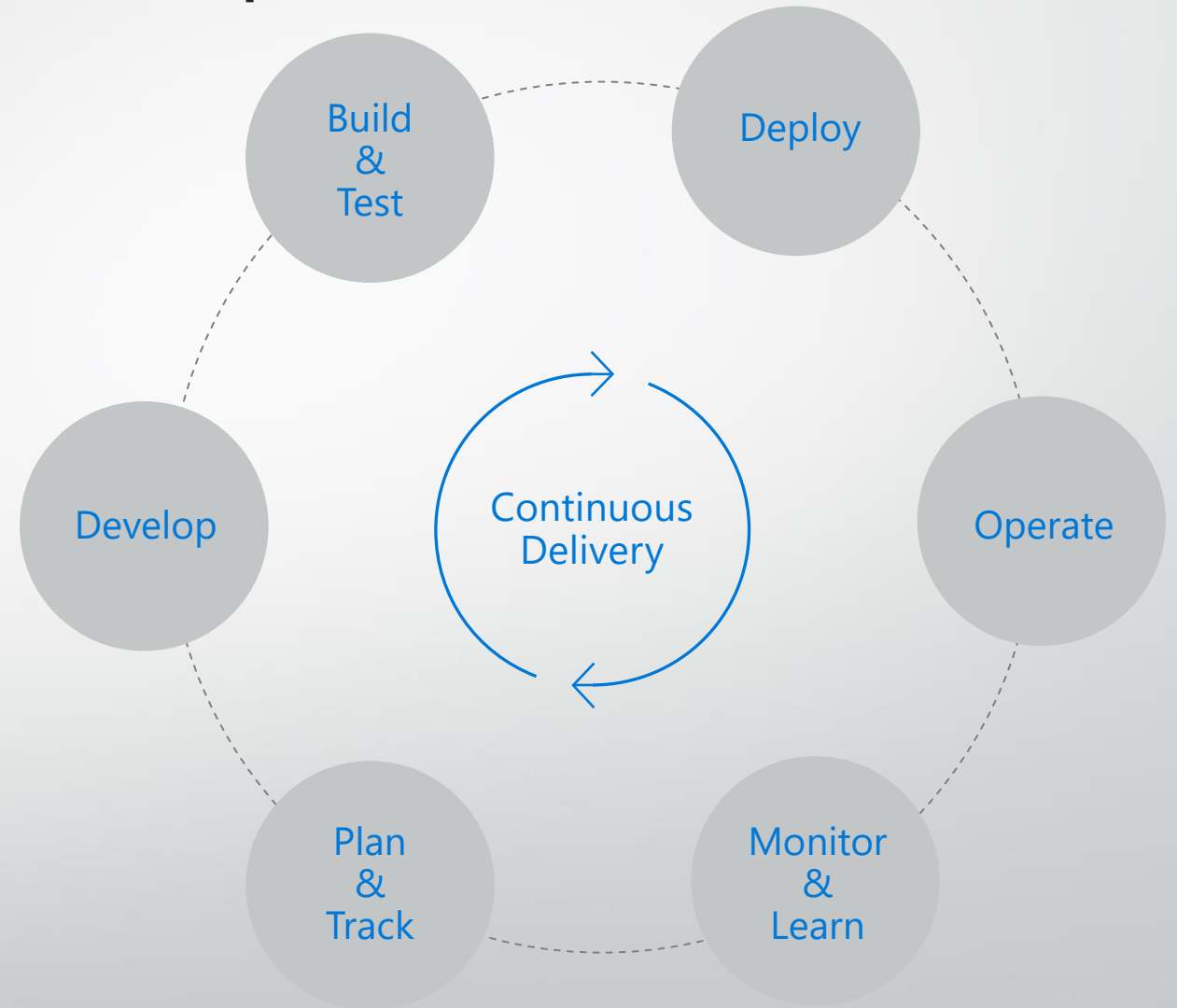
Andrey Vykhodtsev

# What is DevOps?

People. Process. Products.



DevOps is the union of **people**, **process**, and **products** to enable continuous delivery of value to your end users. ”



# Azure DevOps



Azure Boards

Deliver value to your users faster using proven agile tools to plan, track, and discuss work across your teams.



Azure Test Plans

Test and ship with confidence using manual and exploratory testing tools.



Azure Pipelines

Build, test, and deploy with CI/CD that works with any language, platform, and cloud. Connect to GitHub or any other Git provider and deploy continuously.



Azure Artifacts

Create, host, and share packages with your team, and add artifacts to your CI/CD pipelines with a single click.



Azure Repos

Get unlimited, cloud-hosted private Git repos and collaborate to build better code with pull requests and advanced file management.



<https://azure.com/devops>

Print cards from boards,  
queries and backlog

Print whole boards  
or individual work items

The screenshot displays the Agile Cards interface. On the left, a sidebar shows a hierarchy: Backlogs > Queries > Features > Stories > Current > Iteration 1. The main area is divided into two sections. The top section, labeled 'Backlog', shows a list of work items. The bottom section, labeled 'Board', shows a Kanban board with four columns: New, Active, Resolved, and Closed. Each column contains several work items, each represented by a card with a title, a person's name, and a state. The 'Active' column is highlighted with a blue line, and a blue dot is placed on the 'Print work items from Visual Studio Team Services' card. A blue line also connects the 'Print cards from boards, queries and backlog' text to the 'Backlog' and 'Board' tabs. Another blue line connects the 'Print whole boards or individual work items' text to the 'Active' column.

Backlog Board Capacity

Group by Stories Person All [ ] [ ] [ ] [ ]

New Active Resolved Closed

Get rid of messy sticky notes  
Magda Z.  
State New

As Scrum Master I want physical boards  
Bartek Gatz  
State New

Let's have a new standup meeting gathering place  
Marek Went  
State New

Print work items from Visual Studio Team Services  
Bartek Gatz

Put the cards on the board  
Bartek Gatz

Make sure we have some pins  
Bartek Gatz

Assign a different pin color to each of the team members  
Magda Z.

The board needs to be moved to less crowded place  
Marek Went

Store tasks in Visual Studio Team Services  
Magda Z.

Buy the cork board  
Bartek Gatz

Hang the board to the wall  
Marek Went

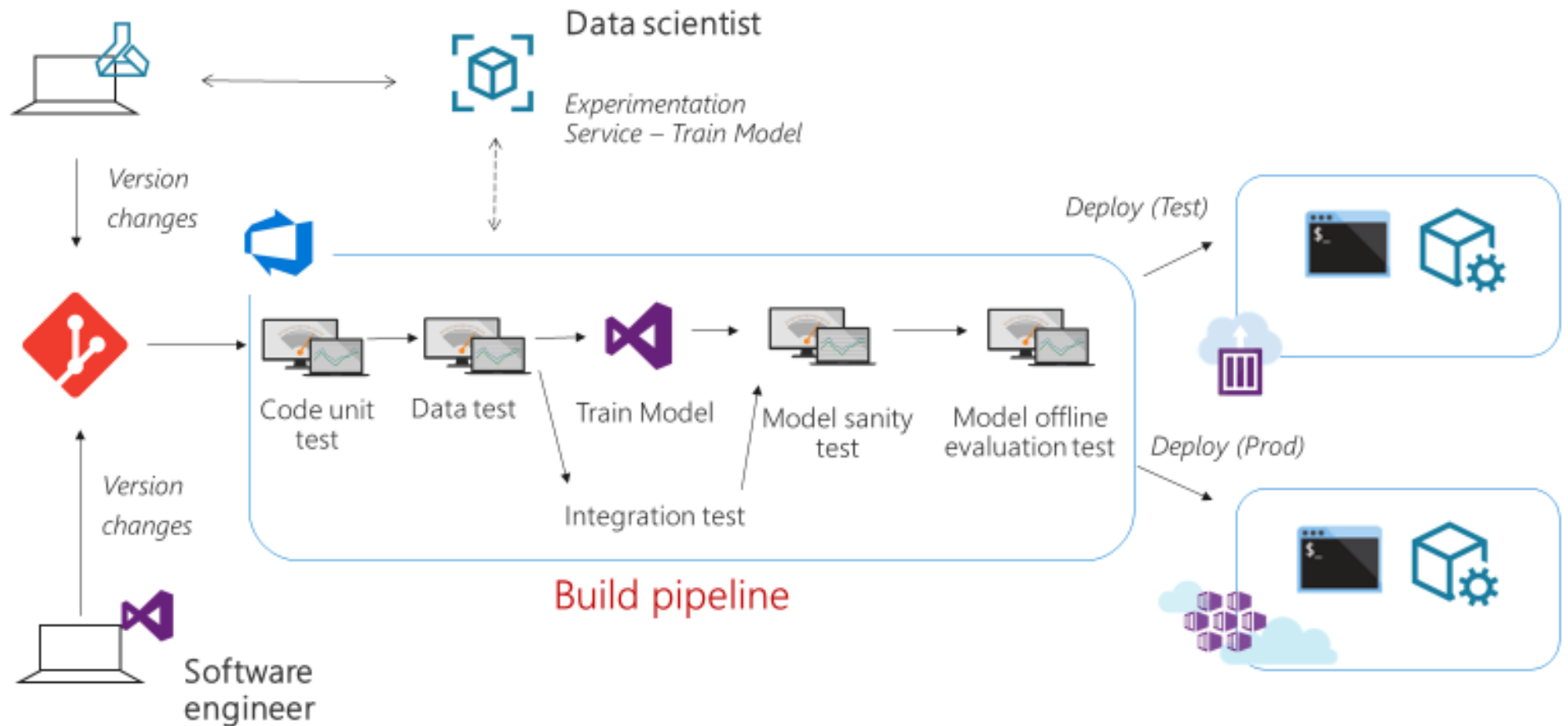
Inform the team of the new standup location  
Magda Z.

Create calendar event for standups with new location  
Bartek Gatz

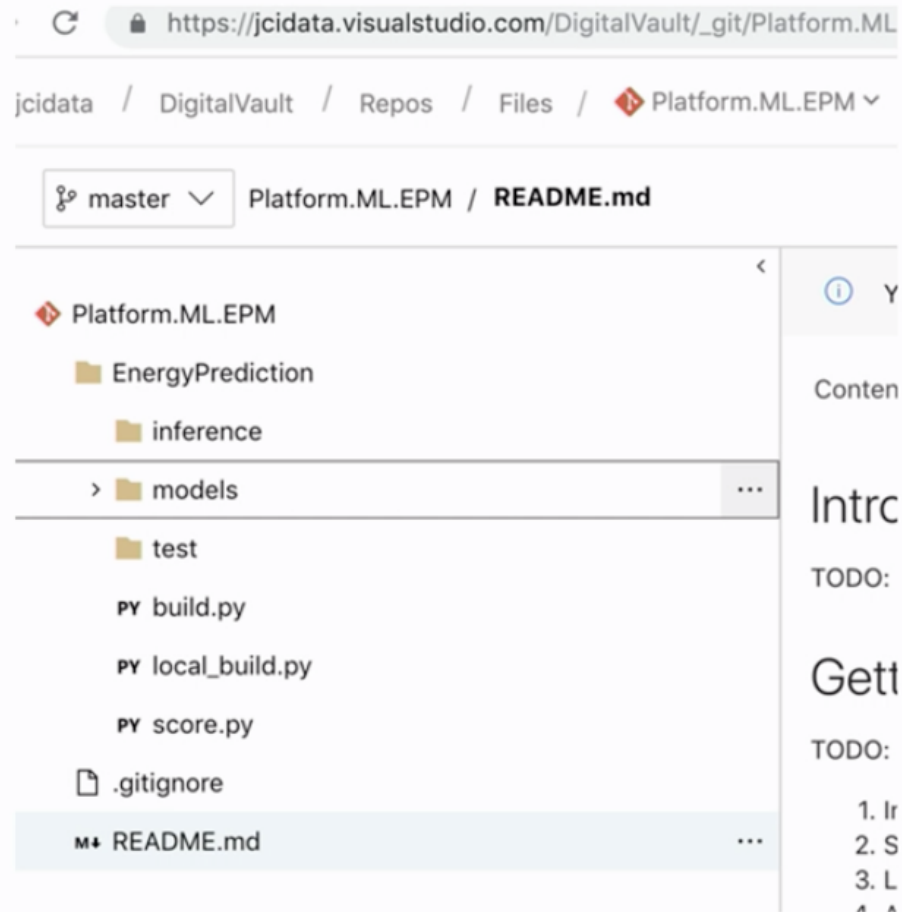
Announce the new standup time  
Magda Z.

# Agile Cards (VS Marketplace)

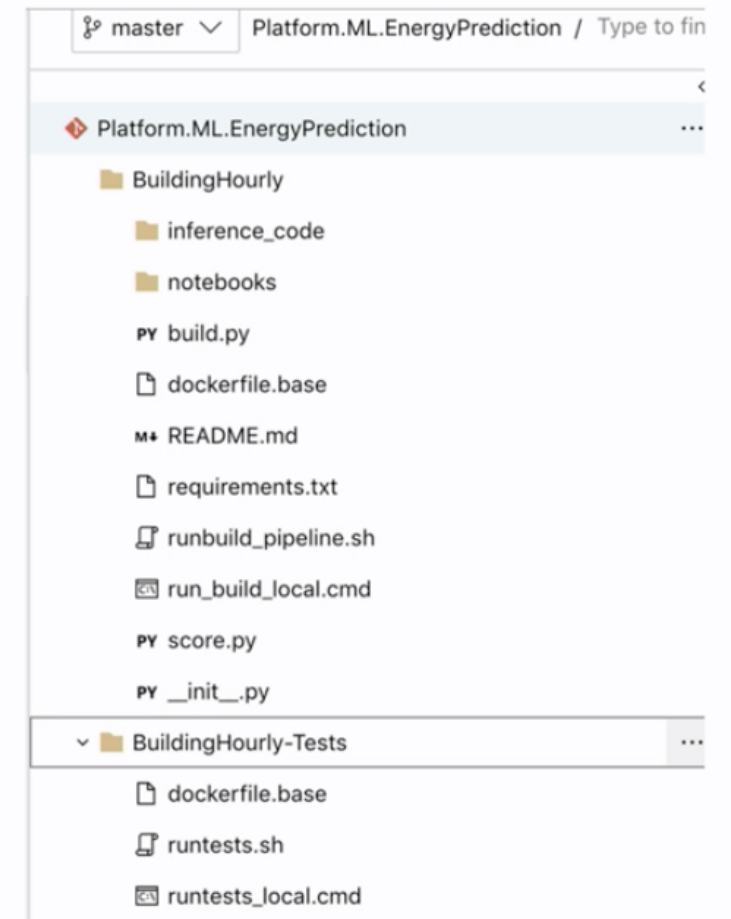
# Integrated Pipeline for Data Scientists and Software Engineers



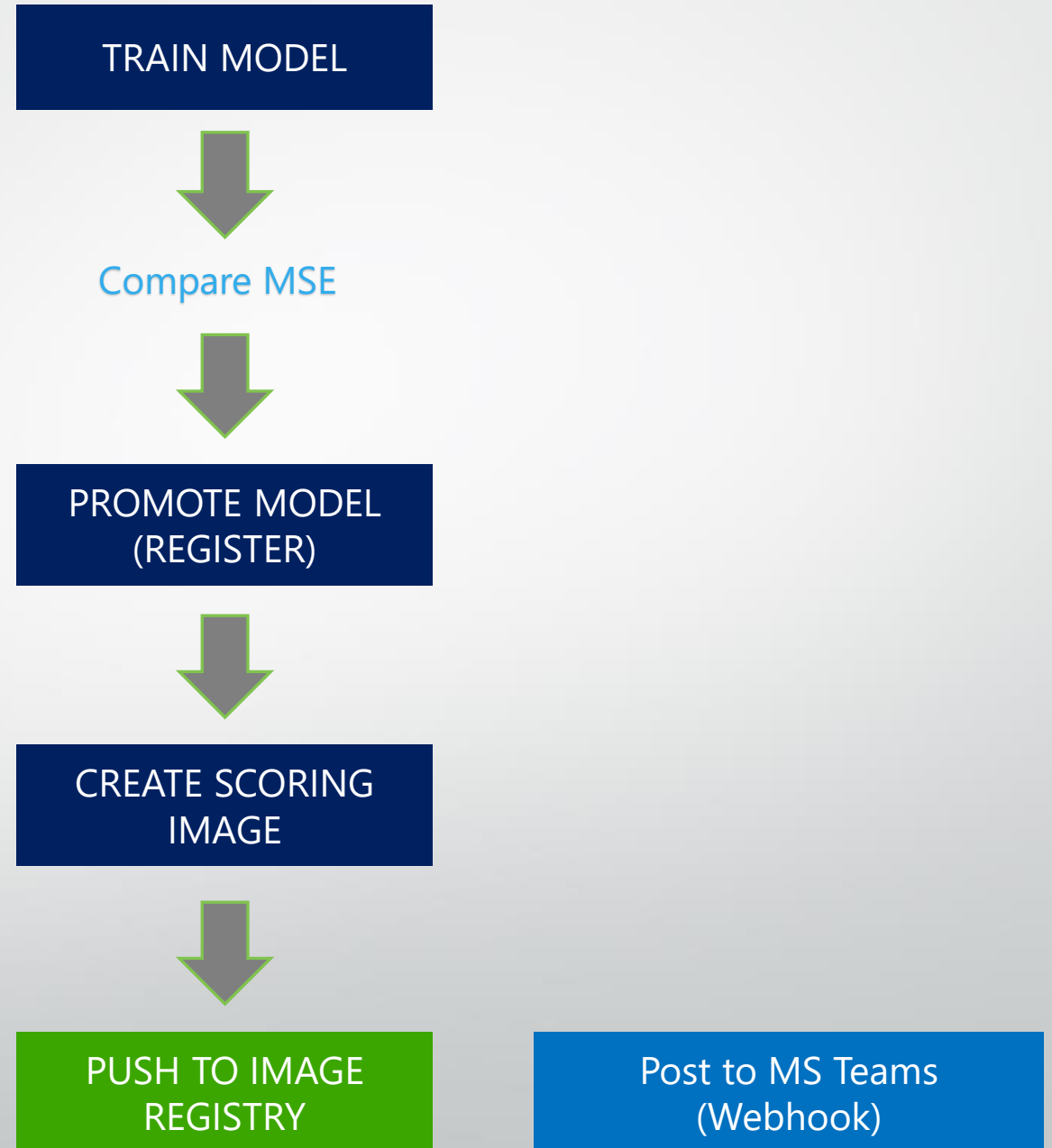
The "old way" ...  
...using git to be our "manager of models"



The "new way" ...  
...Full CI/CD! ZOMG!  
Git is no longer the model manager



# BUILD PIPELINE



# RELEASE PIPELINE

QA

if (new\_image)



CREATE AZURE  
CONTAINER  
INSTANCE FROM  
NEW IMAGE



TEST WITH  
VALIDATION SET



Post to MS Teams  
(Webhook)

PROD

UPDATE AKS  
DEPLOYMENT WITH  
NEW IMAGE



TEST WITH  
VALIDATION SET



# Advanced analytics pattern in Azure

Data collection and understanding, modeling, and deployment

