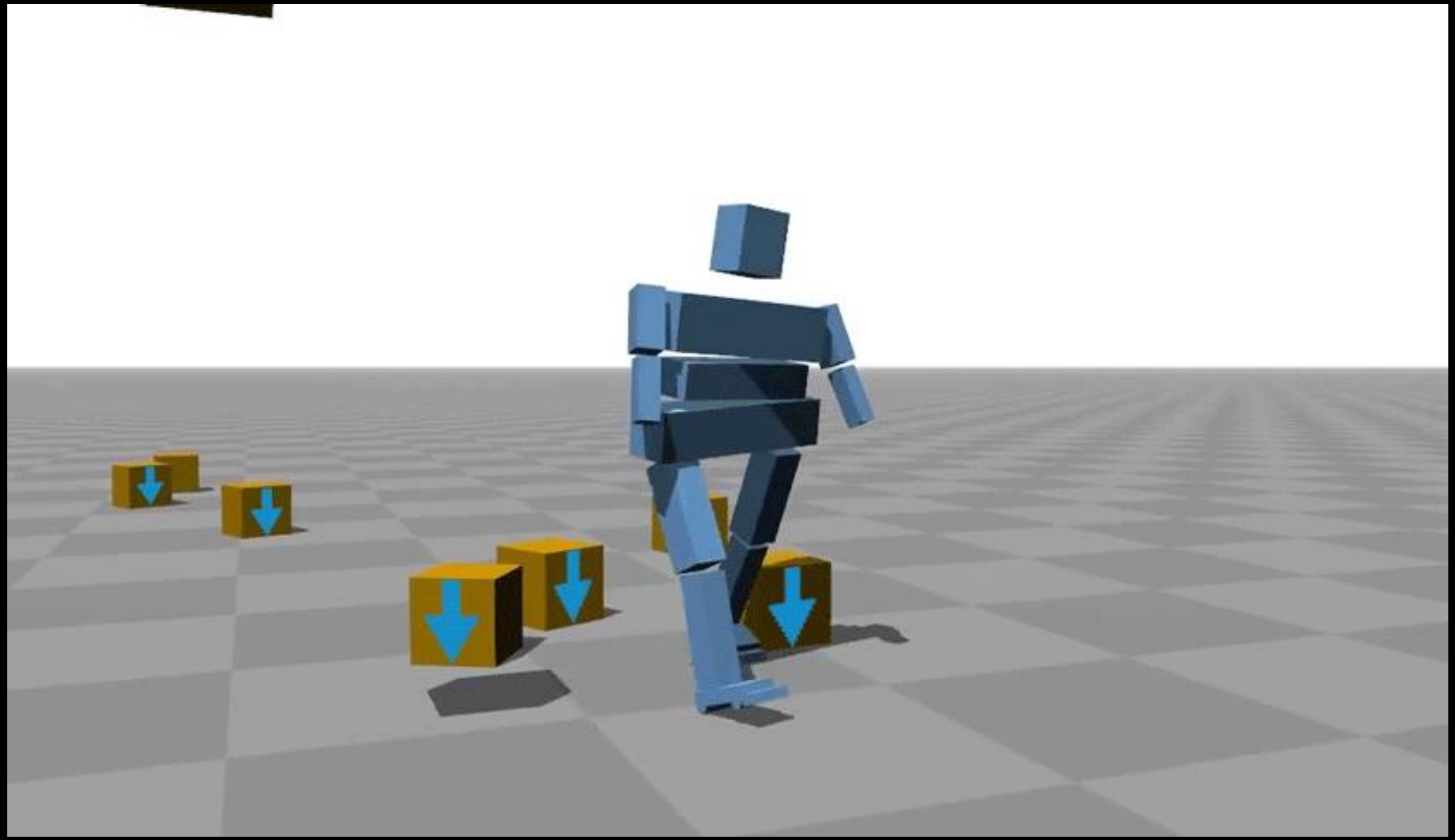


MLOps
CI/CD for Machine Learning

SASHA ROSENBAUM

Sasha Rosenbaum

Sr. Program Manager

Microsoft

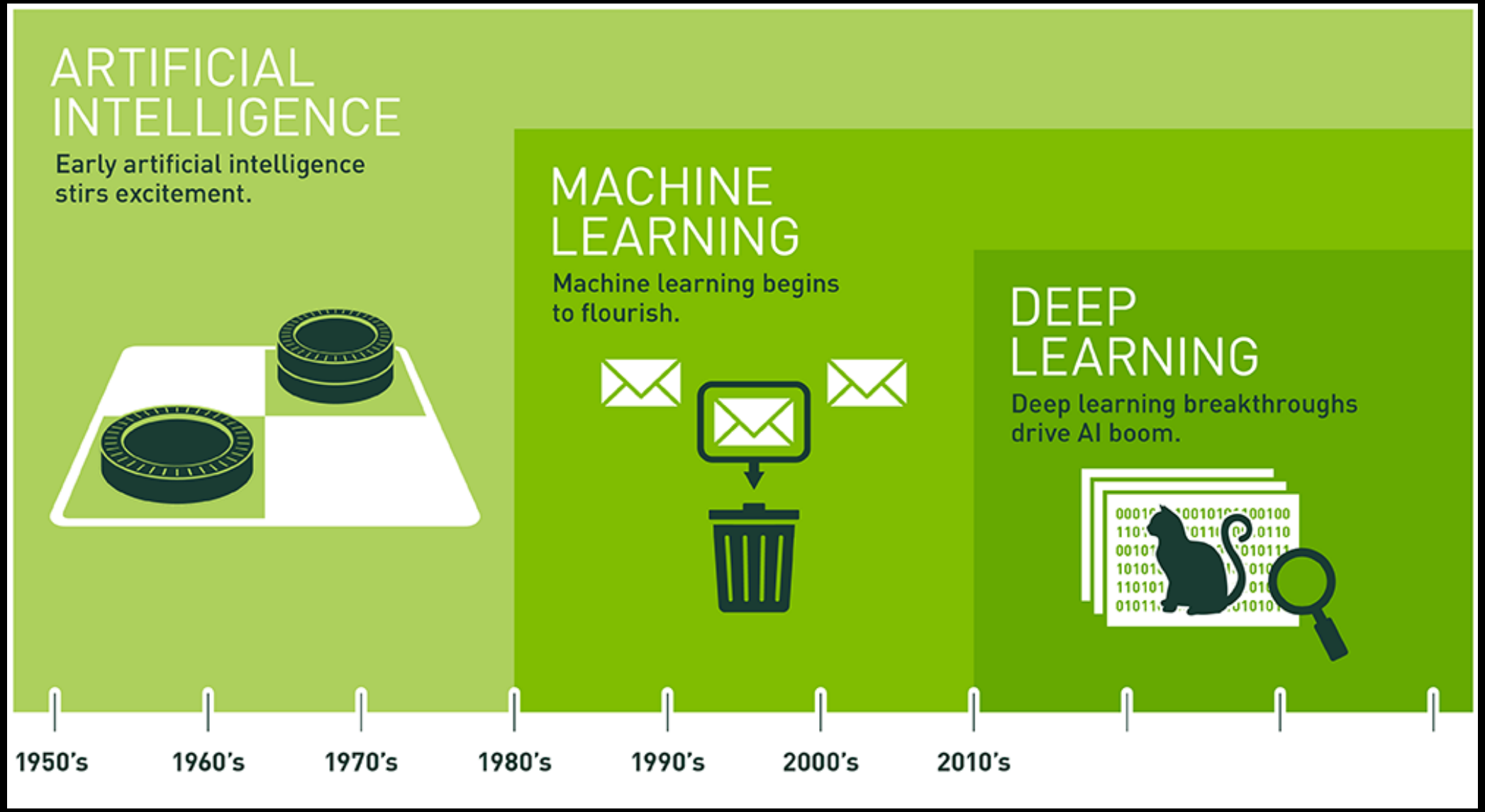
@DivineOps

What is MLOps
and
WHY should you care?

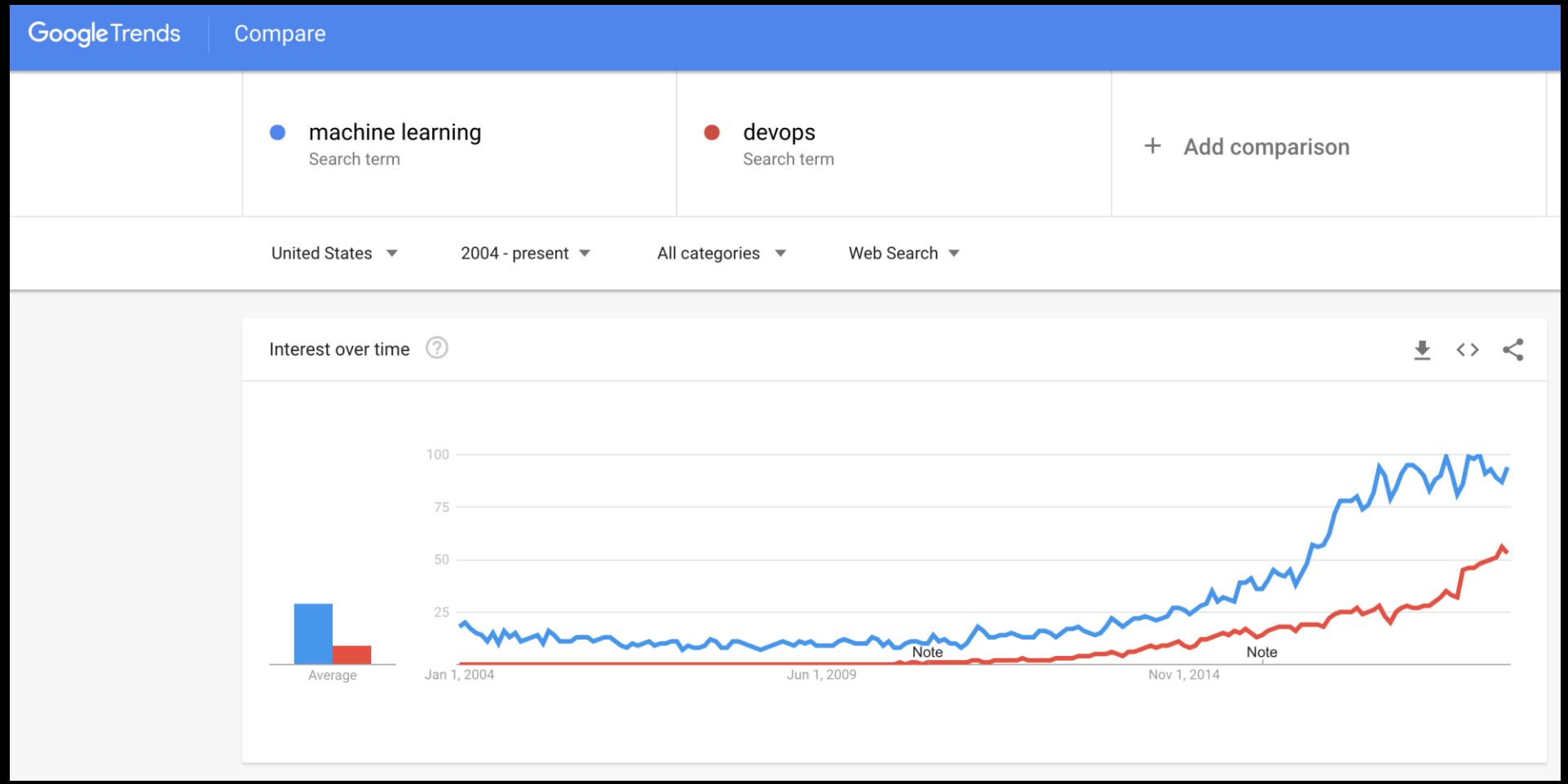
Machine Learning (ML)

Is the science of getting computers to act

Without being explicitly programmed

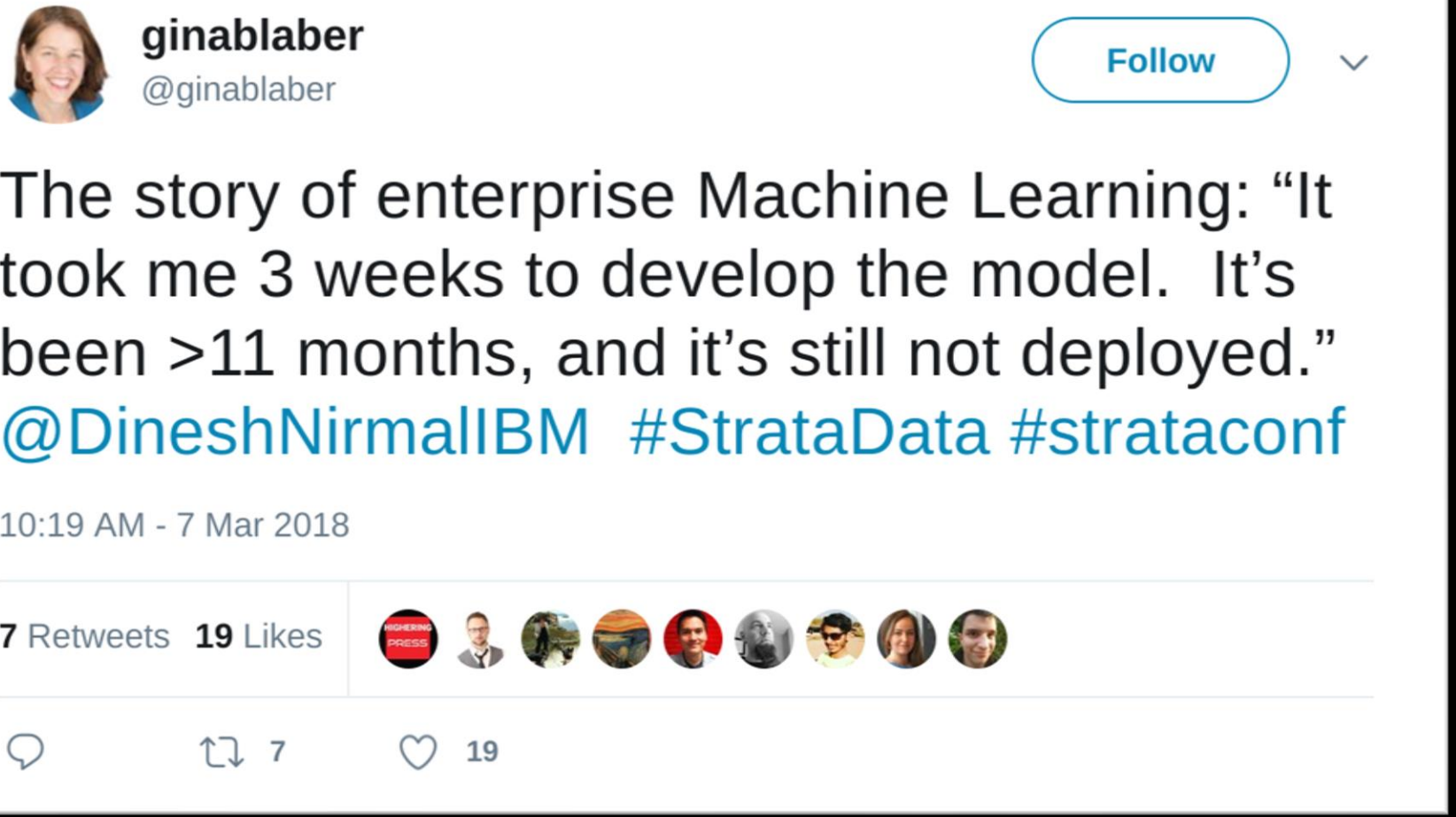

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

Machine Learning vs DevOps Google searches



Python questions on Stack Overflow

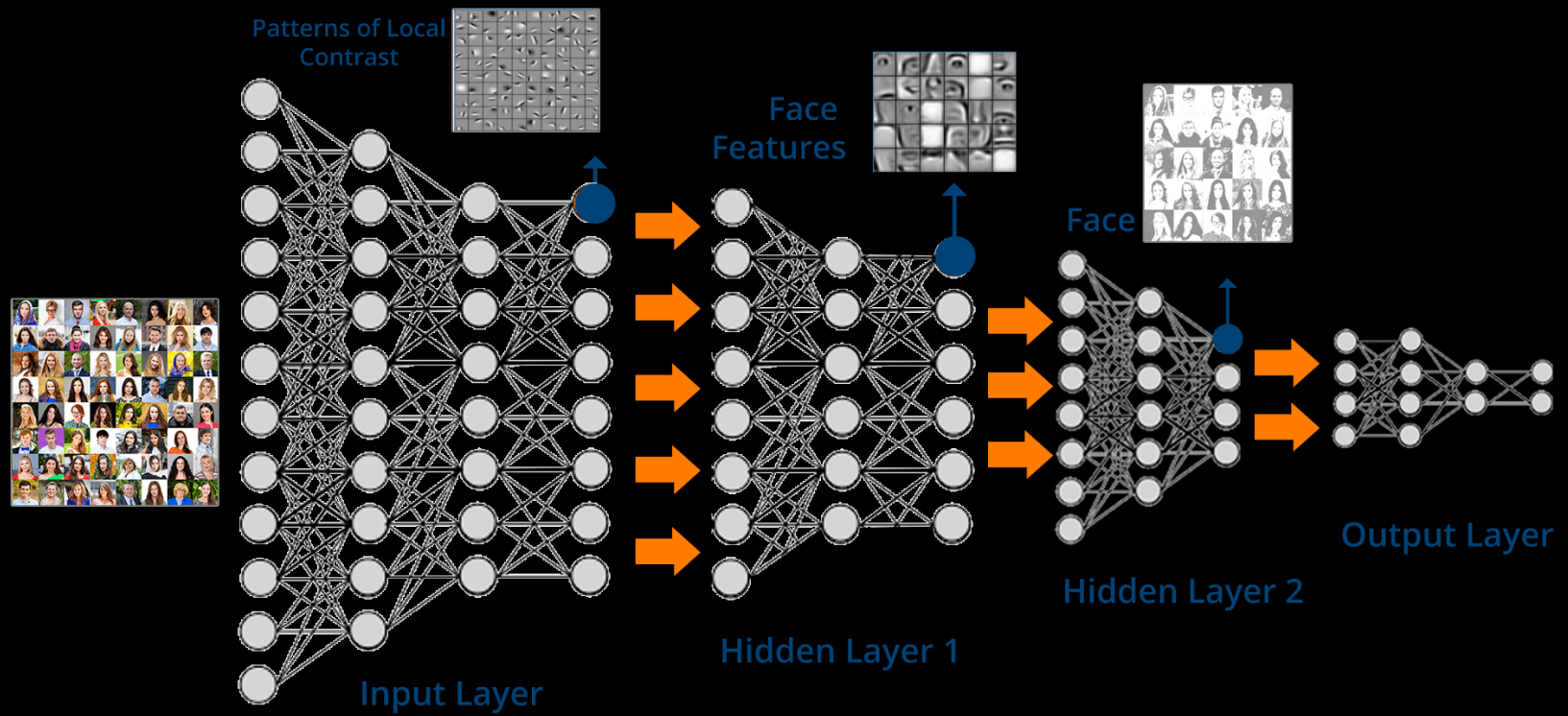
OK, but why should YOU care?


ginablaber
@ginablaber
Follow
The story of enterprise Machine Learning: “It took me 3 weeks to develop the model. It’s been >11 months, and it’s still not deployed.” @DineshNirmalIBM #StrataData #strataconf
10:19 AM - 7 Mar 2018
7 Retweets 19 Likes
7
19

Data Scientists just want to Data Science

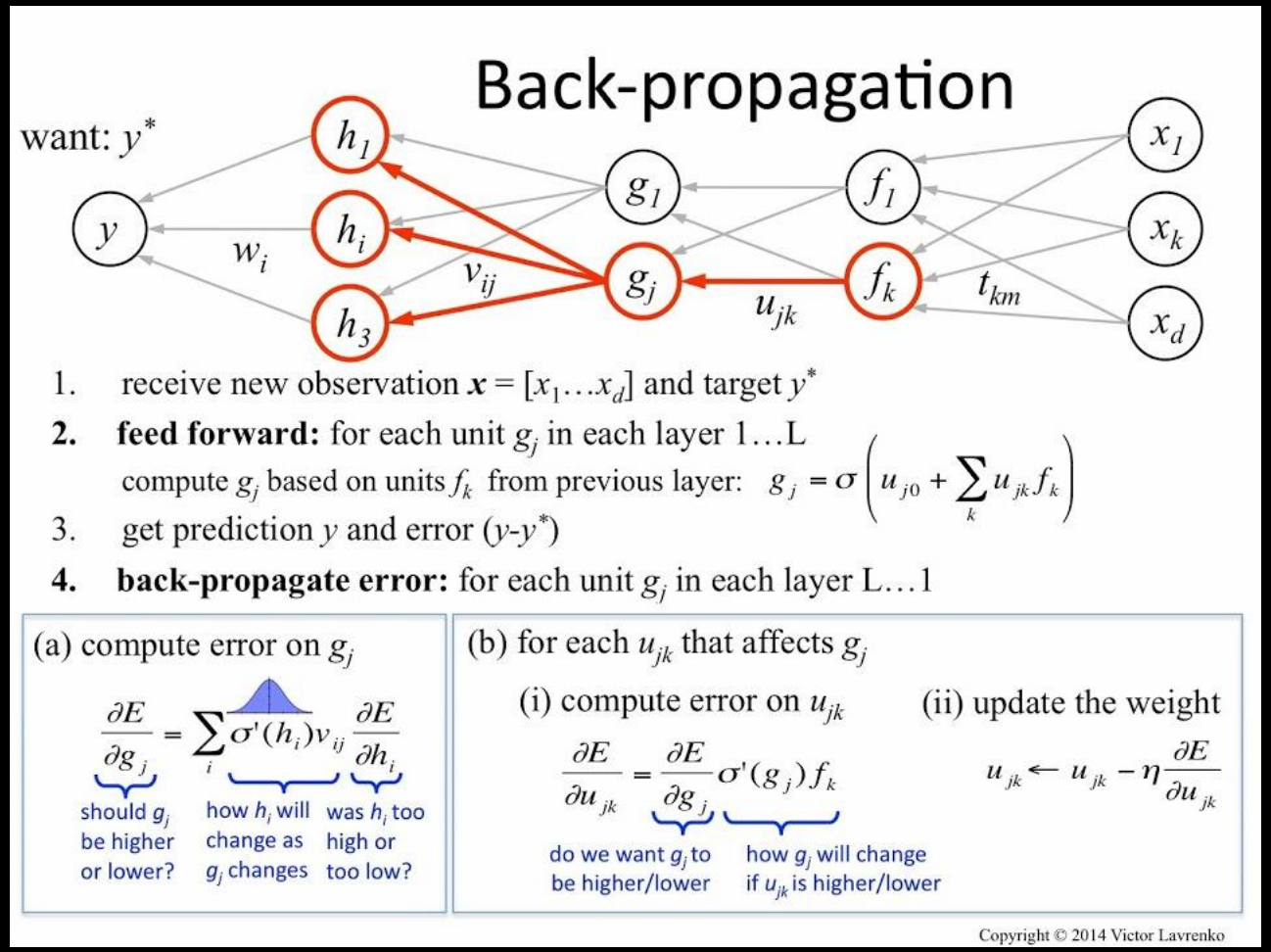
Deep Learning

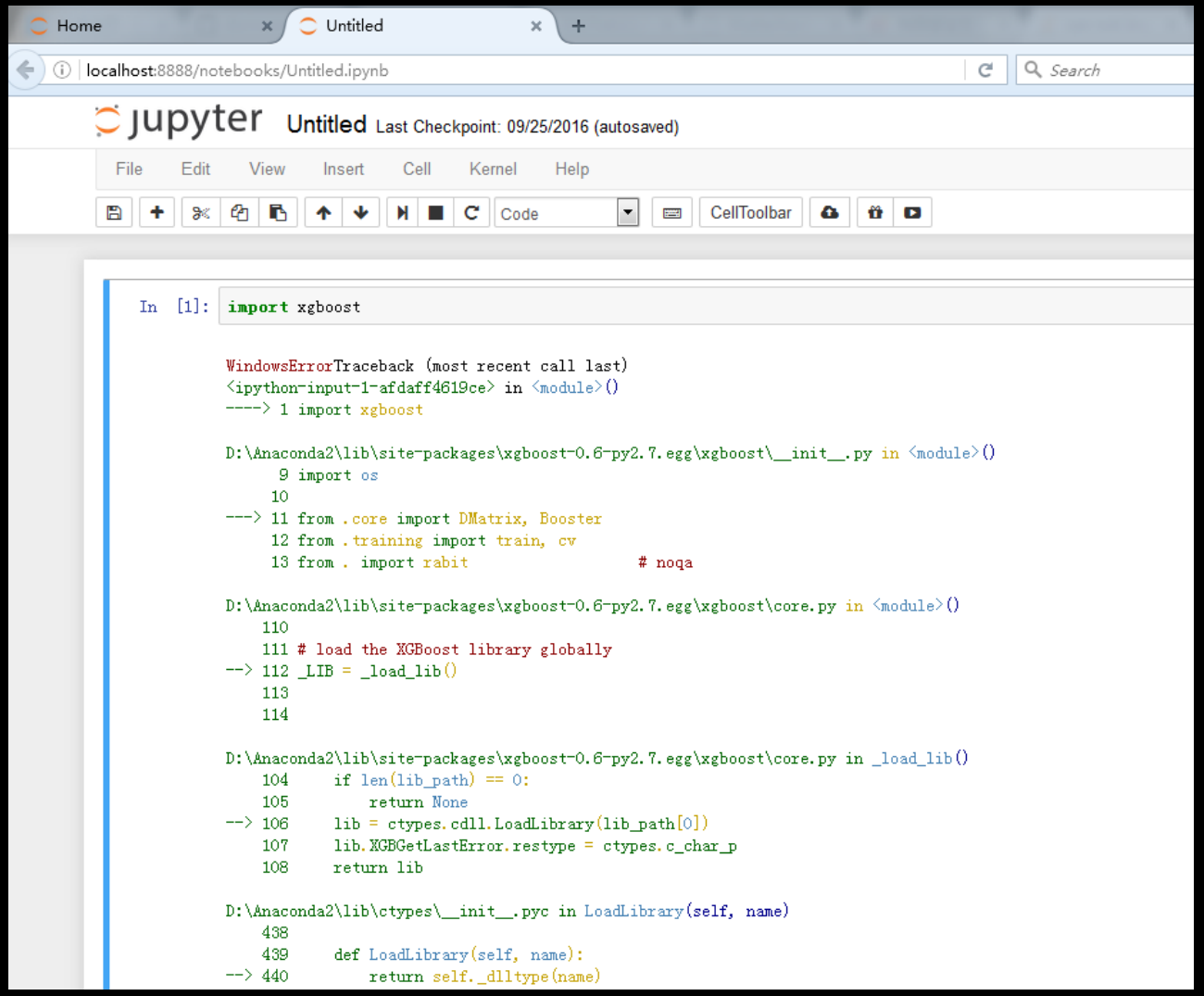
Some ML training algorithms are complex



Deep Learning - Backpropagation

Some ML training algorithms are complex





Typical data scientist work environment

We've got the notebook into source control!

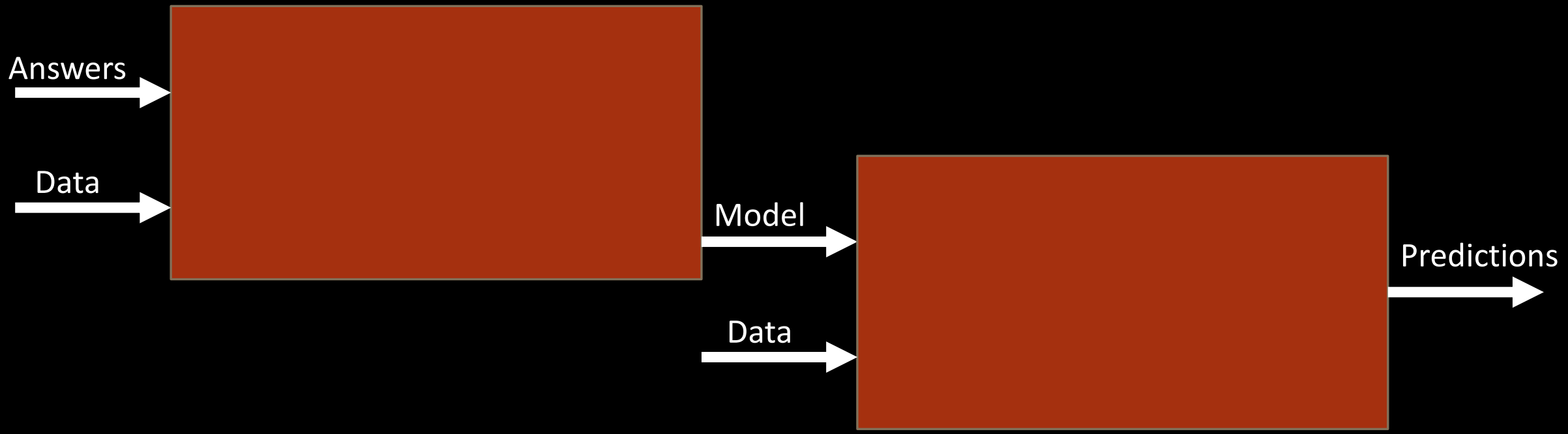
Programming

Machine Learning

Machine Learning

Machine Learning

Machine Learning



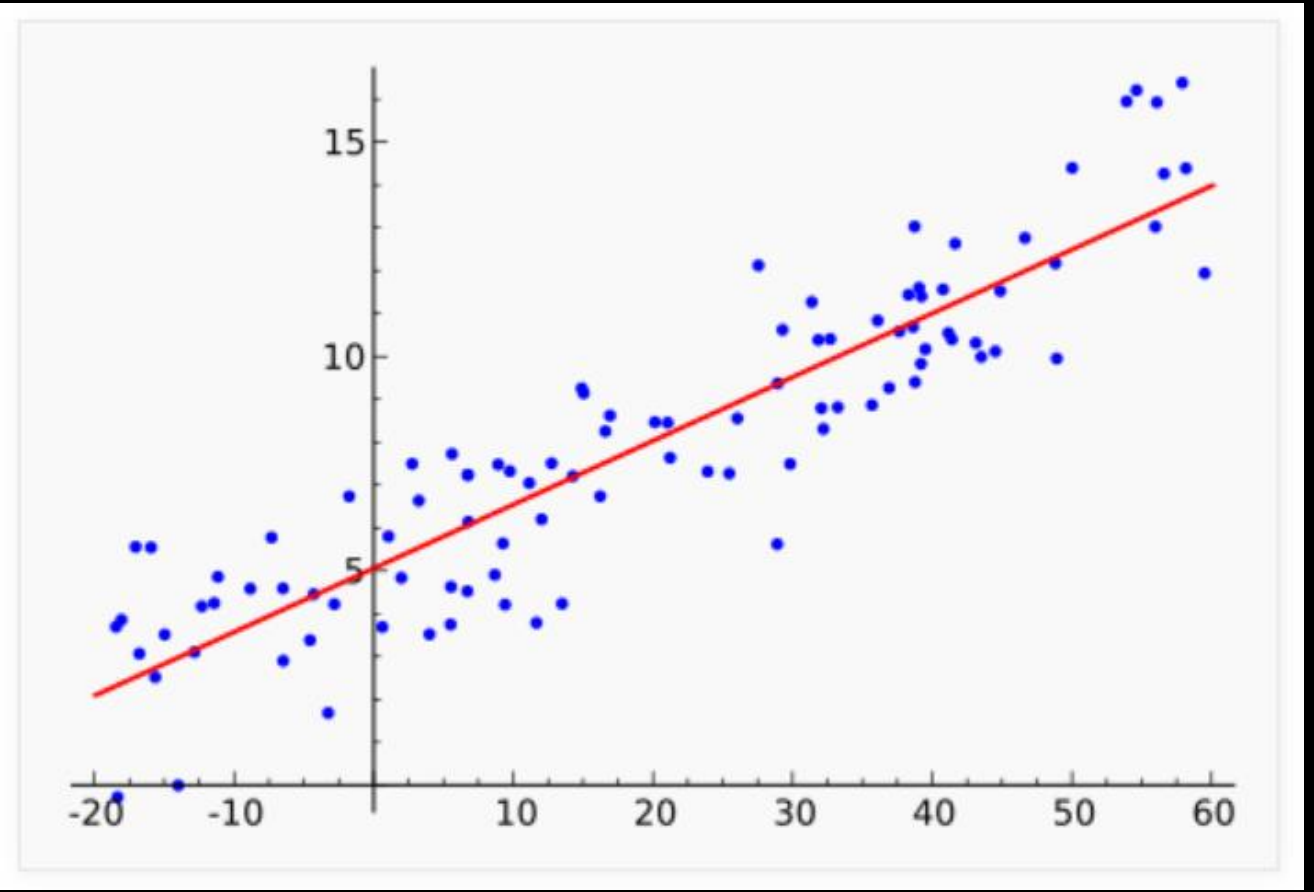
How do we put the model in production?

What is an ML model?

Linear Regression – Housing Prices

The training finds a and b such that

$$Y = a+bX+\epsilon$$



Deep Learning

The input and output may be vectors $\bar{X}, \bar{Y}$

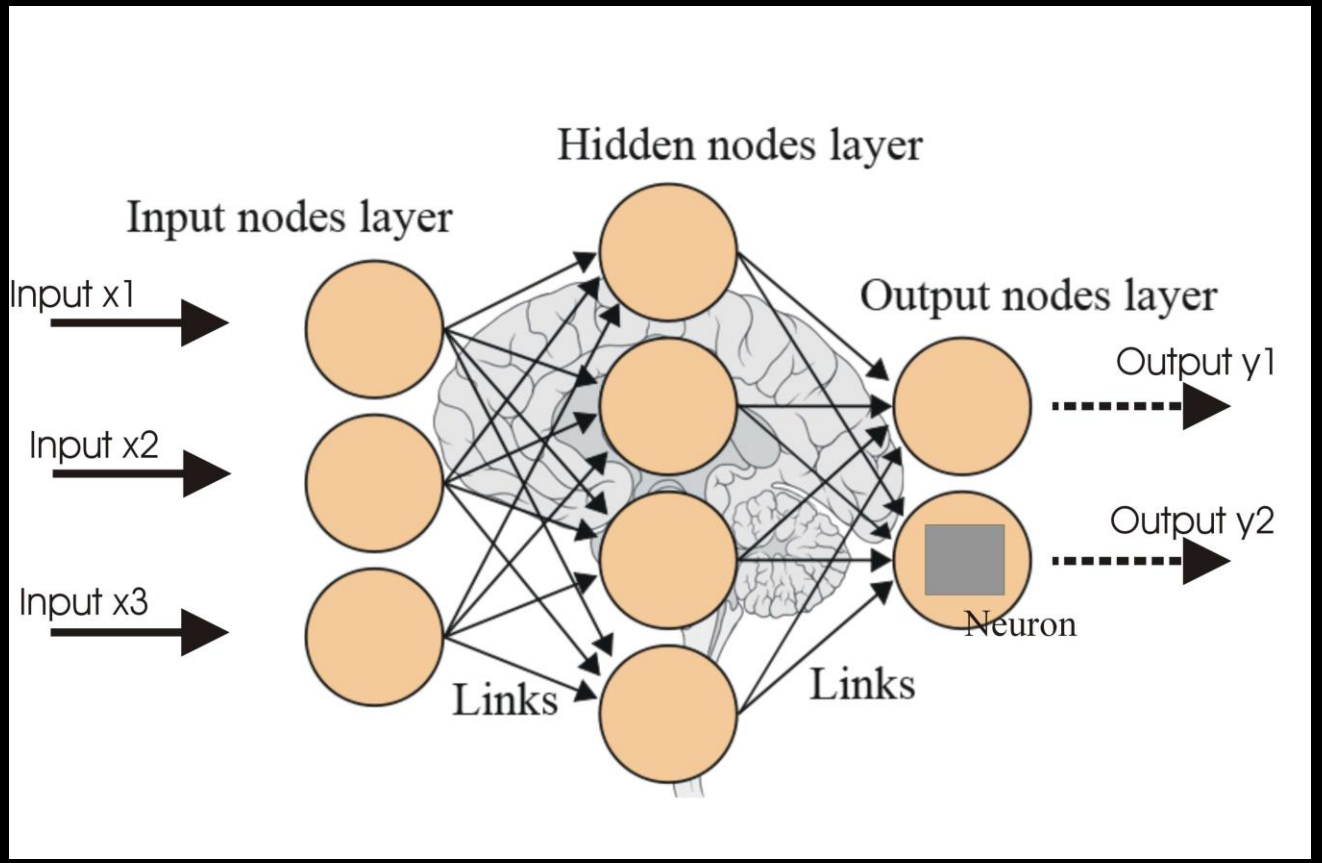


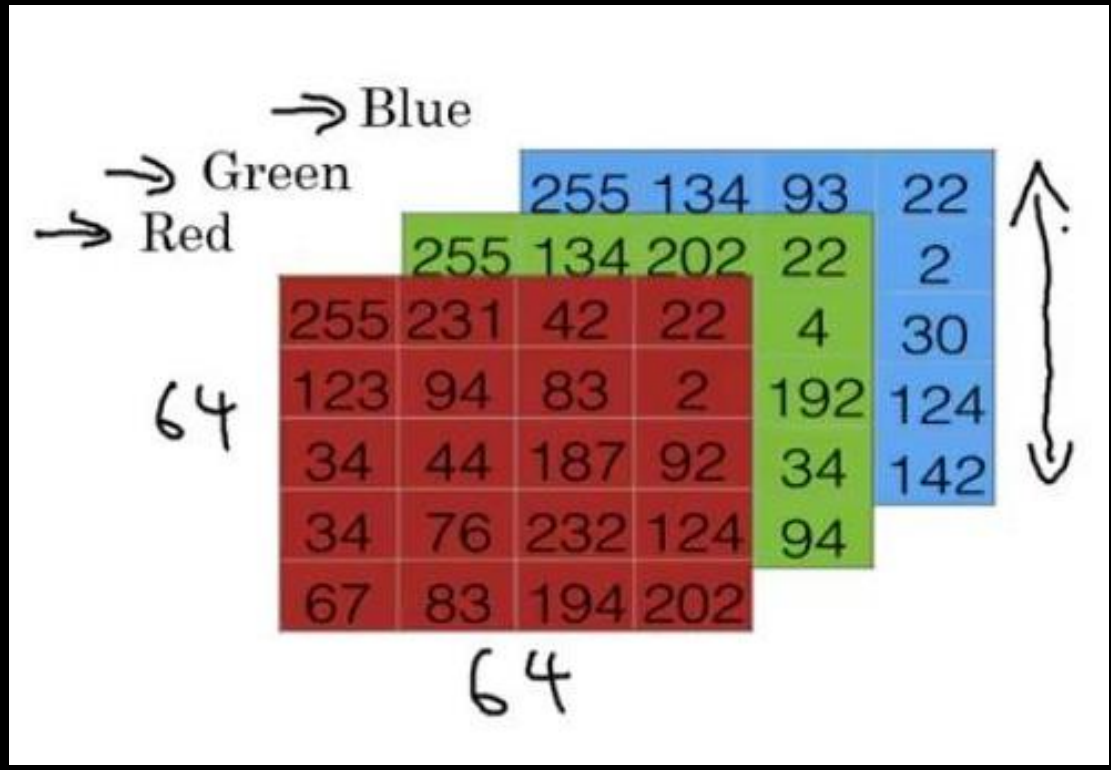
Image Classification

Binary Classification

1 (cat) vs 0 (non cat)

y

To store an image your computer stores three separate matrices corresponding to the red, green, and blue color channels of this image.



Andrew Ng

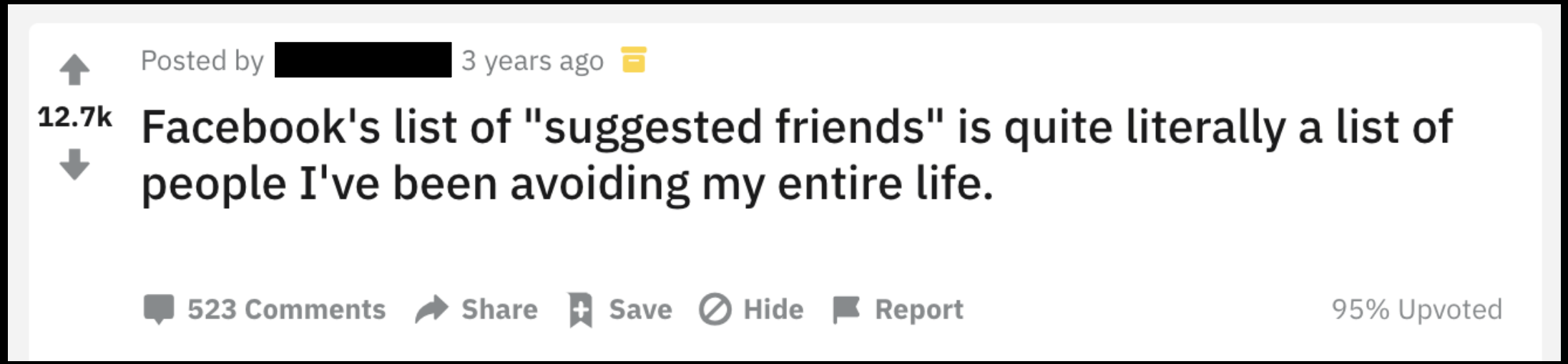
ML Model

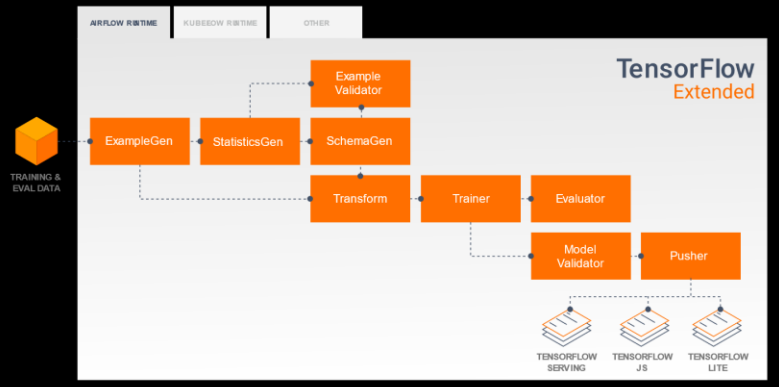
A definition of the **mathematical formula** with a number of parameters that are learned from the data

Isn’t this just an API endpoint?!

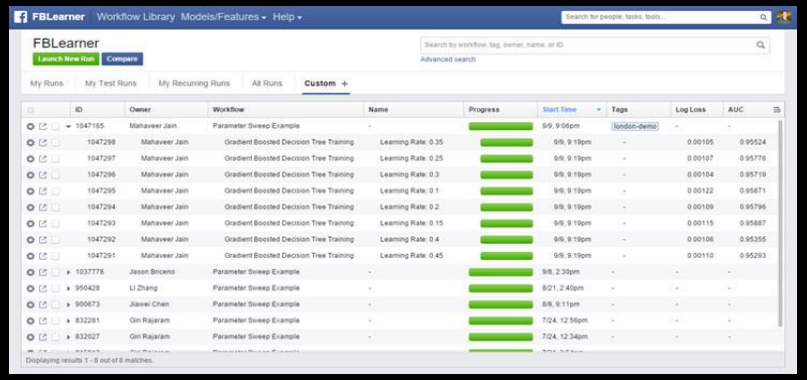
Do models really change that often?

Models must be improved continuously

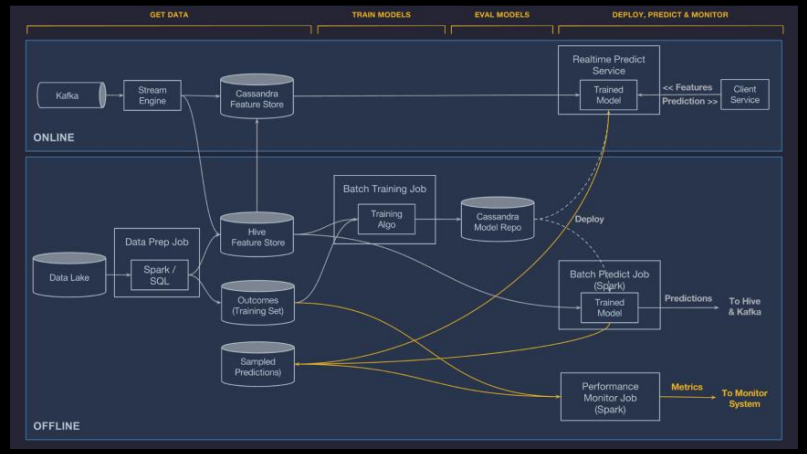




TensorFlow Extended



FBLearner Flow



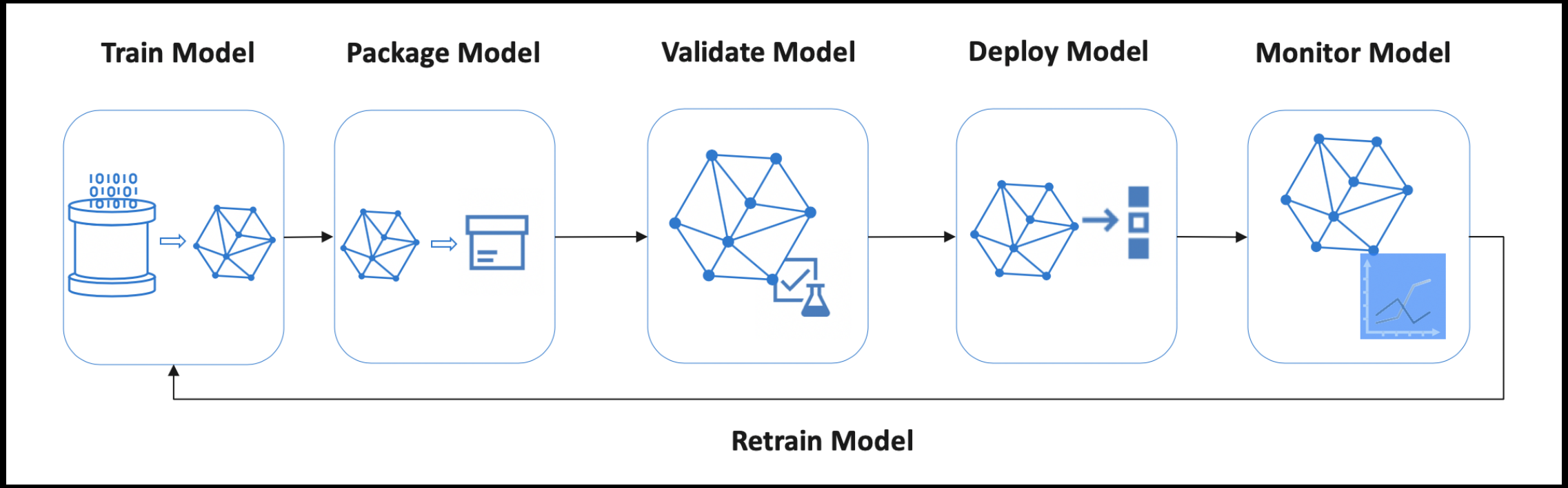
Uber's Michelangelo

Microsoft Aether

But I don’t work at a big company with thousands of ML engineers!

How do we iterate?

Machine Learning Lifecycle



Data Scientist

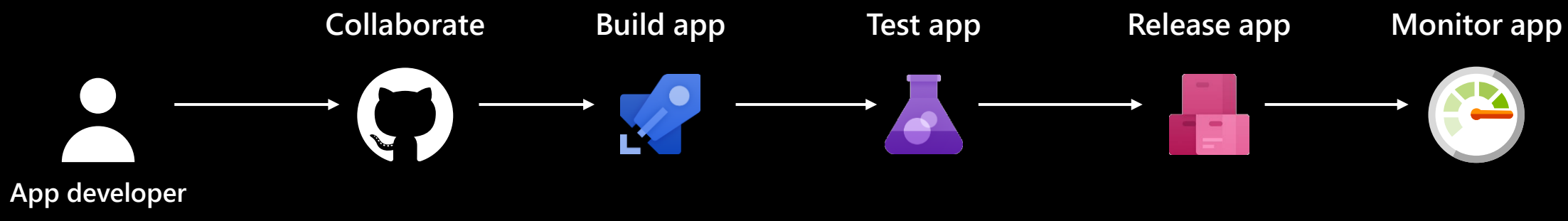
- Quick iteration
- Versioning
- Reuse
- Great tools
- Ease of management
- Unlimited scale
- Eliminating drift

Friends?

DevOps/SRE

- Quick iteration
- Versioning
- Reuse
- Compliance
- Observability
- Uptime
- Updates

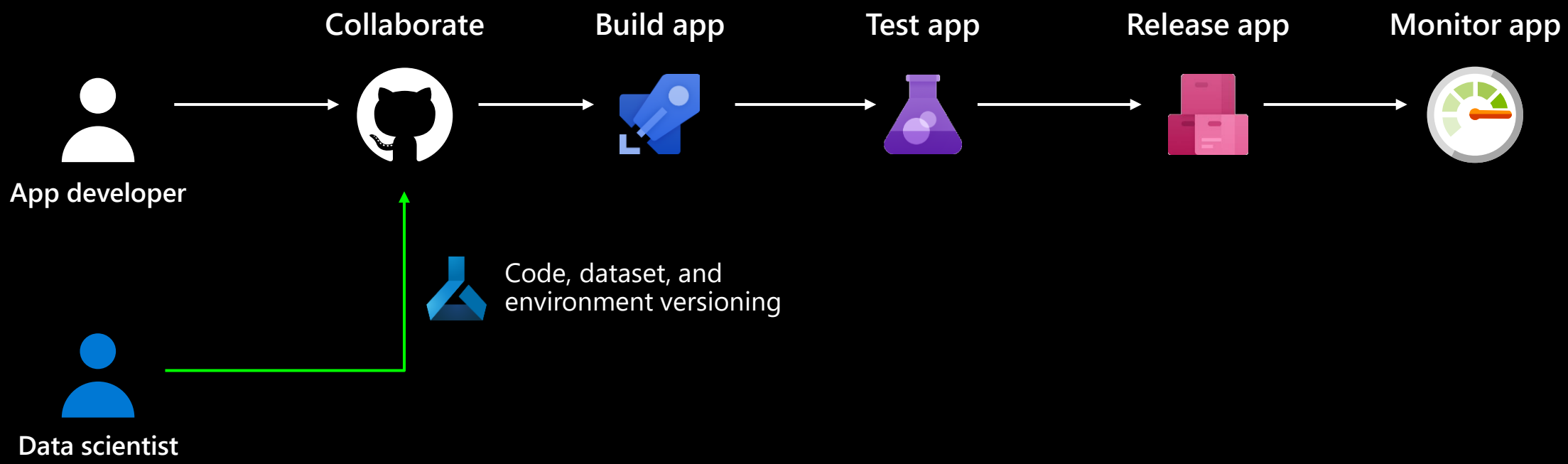
MLOps Workflow



Data scientist

- Model reproducibility
- Model validation
- Model deployment
- Model retraining

MLOps Workflow



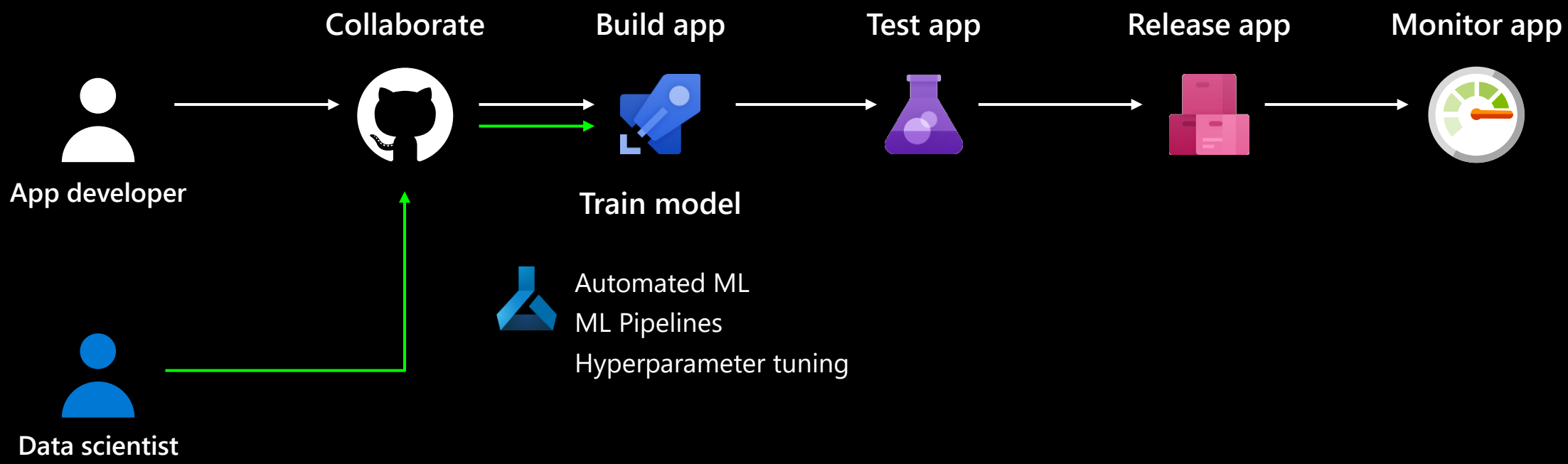
Model reproducibility

Model validation

Model deployment

Model retraining

MLOps Workflow



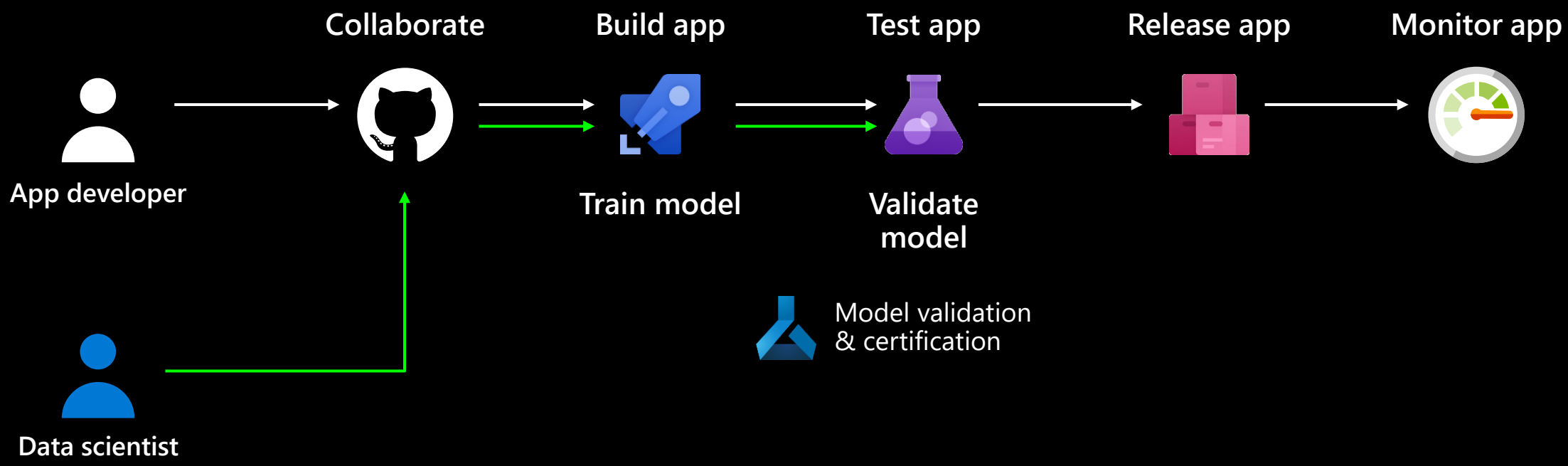
Model reproducibility

Model validation

Model deployment

Model retraining

MLOps Workflow



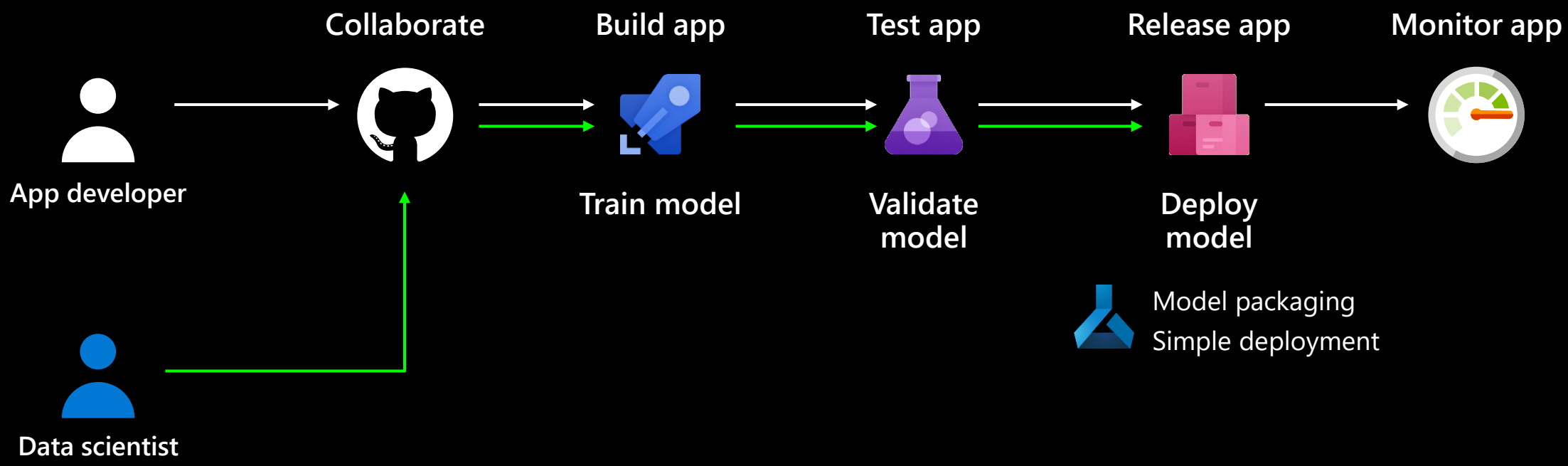
Model reproducibility

Model validation

Model deployment

Model retraining

MLOps Workflow



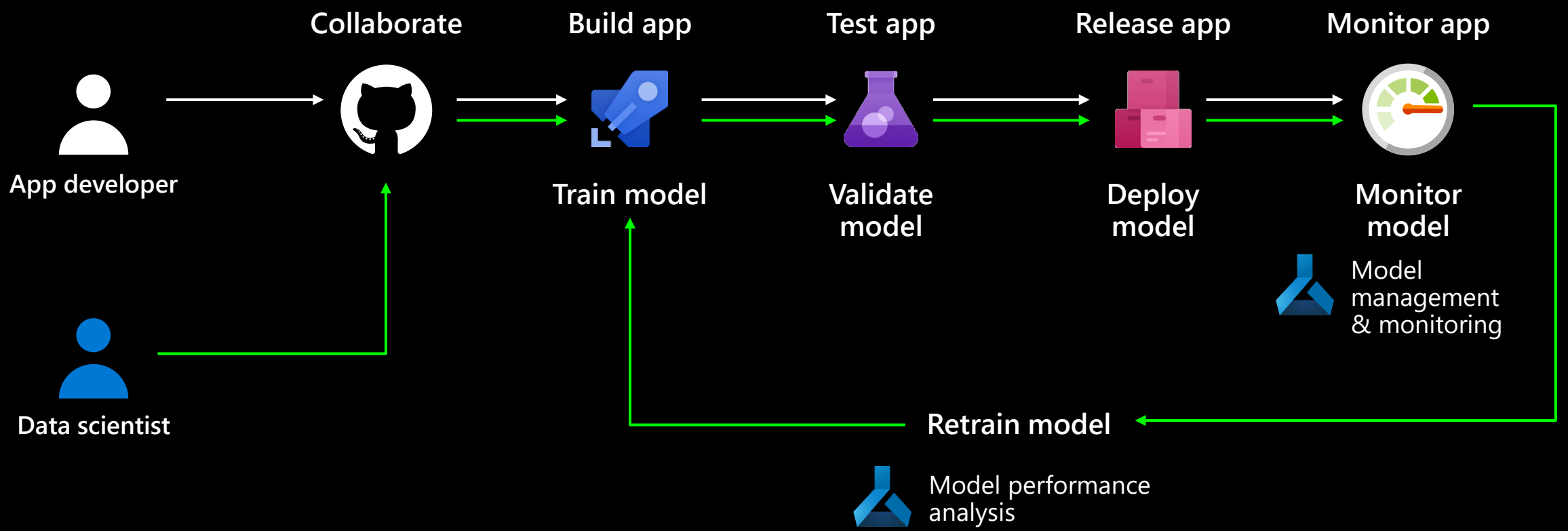
Model reproducibility

Model validation

Model deployment

Model retraining

MLOps Workflow



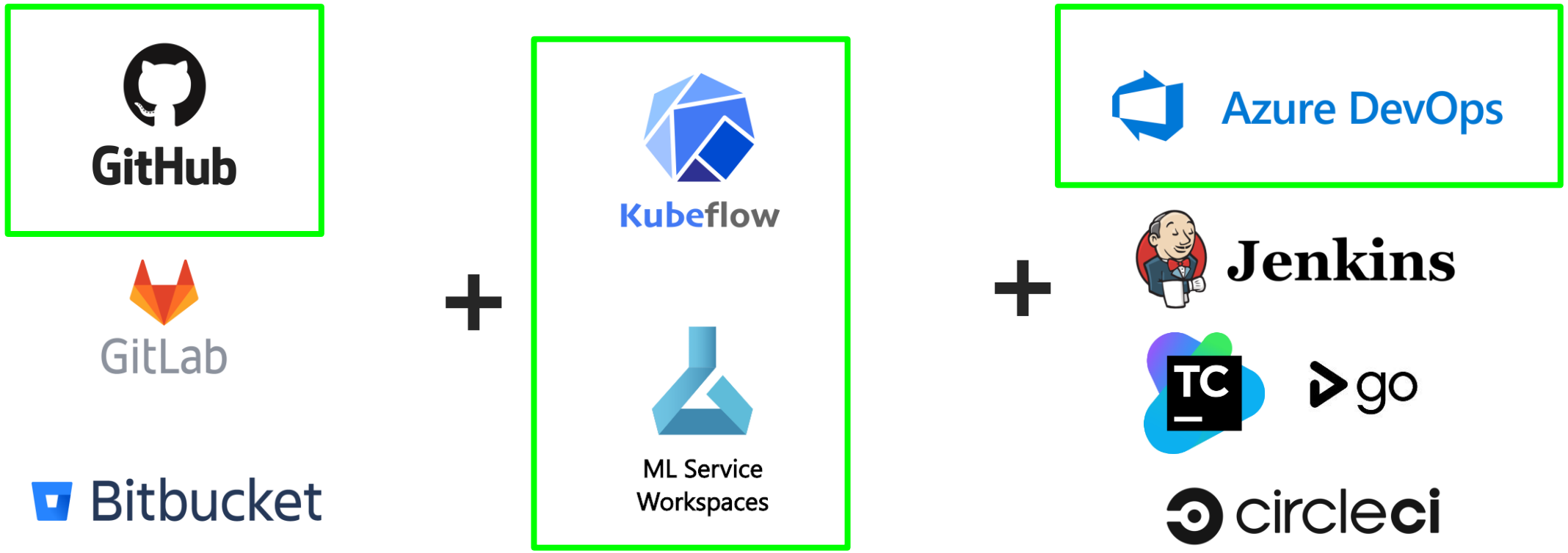
Model reproducibility

Model validation

Model deployment

Model retraining

Build Your Own MLOps Platform

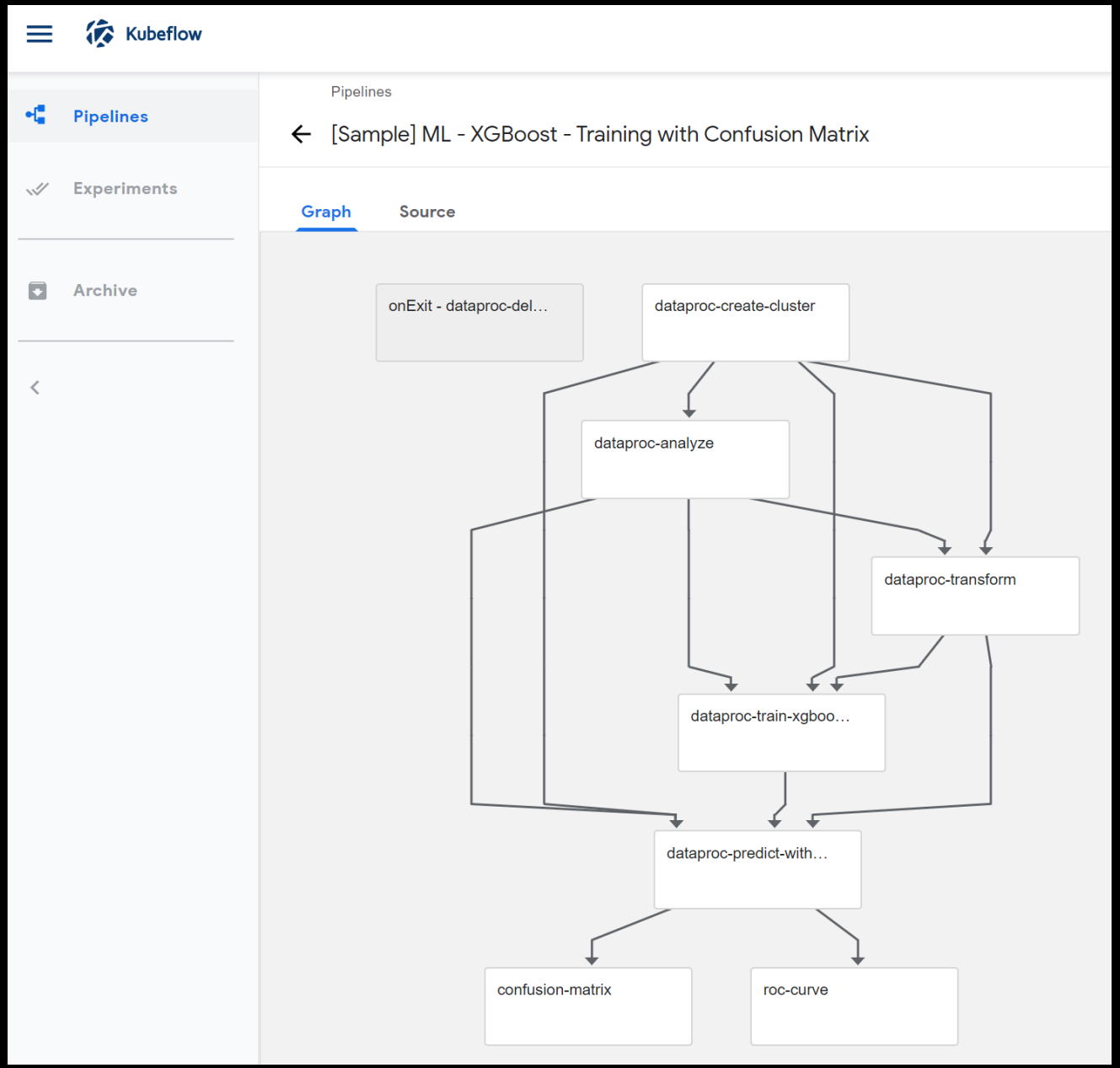


ML Pipeline

A reusable, scaleable ML workflow template

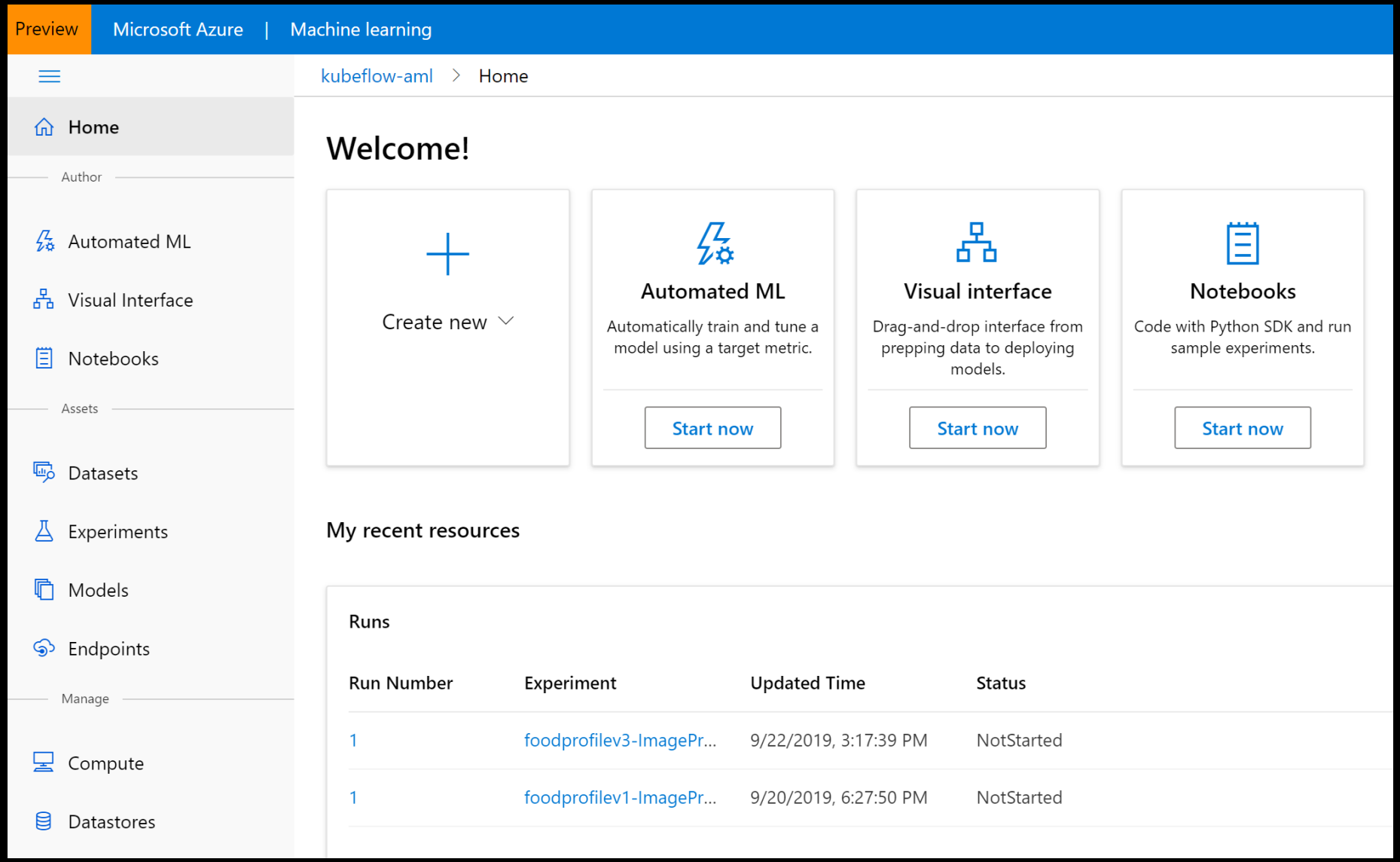
Kubeflow pipeline

A reusable, scalable ML workflow template that runs on containers



Azure ML

- Prep data
- Train
- Test
- Deploy
- Manage



The CI/CD Pipeline

GitHub Repo https://github.com/DivineOps/kubeflow-and-mlops

MLOps with Kubeflow, Azure Machine Learning and Azure Pipelines

This repository provides a sample ML CI/CD pipeline using Kubeflow, Azure ML workspaces and Azure Pipelines.

It requires:

- An Azure Account (A trial account works!)
- An Azure DevOps Organization (The free tier works!)

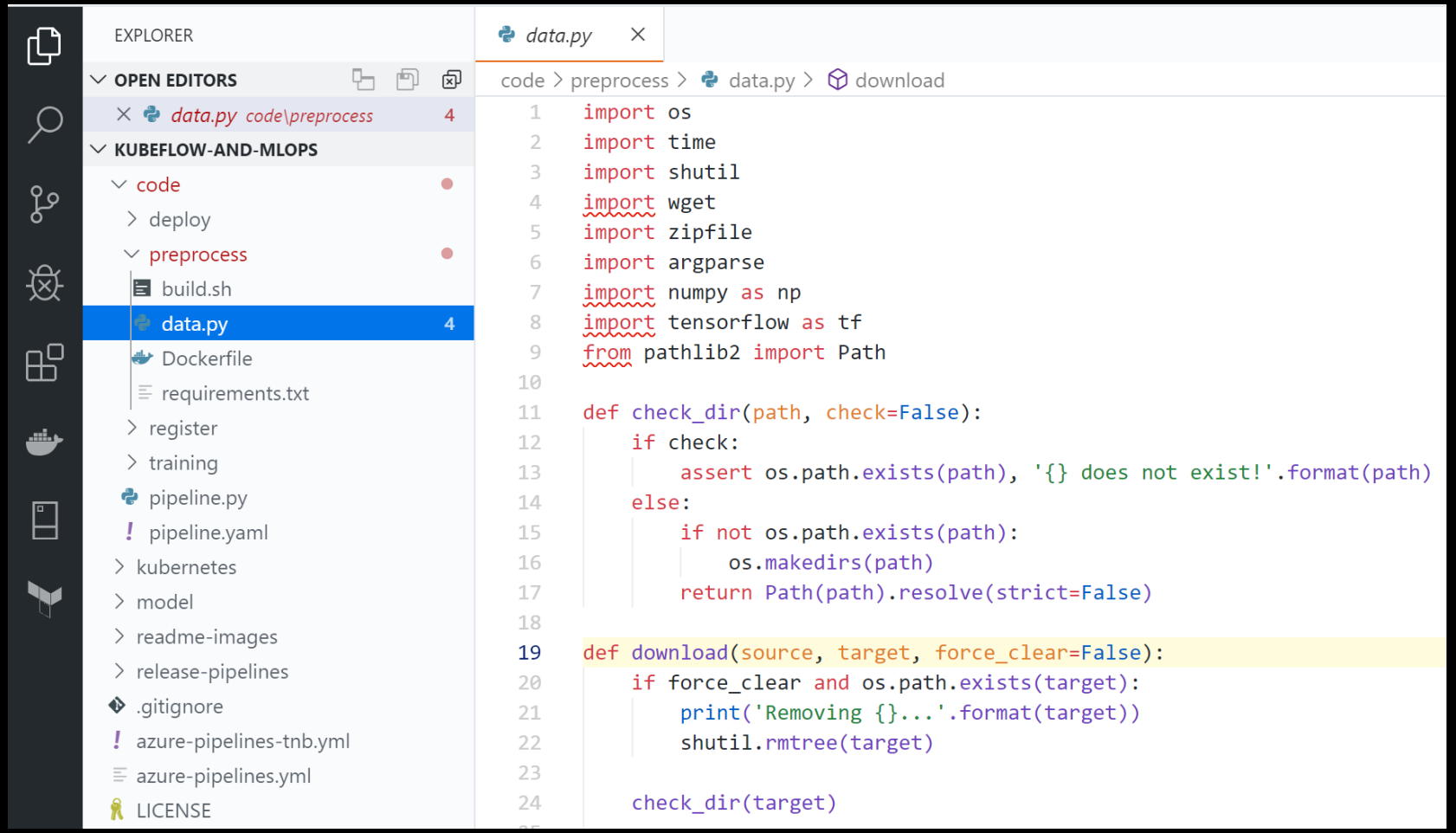
Slides

Link to the slides to the All Things Open version of this talk Slides

Azure Resources

aks-cluste	Kubernetes service
fooddeplo	Container instances
kf-aks	Kubernetes service
kubeflowa	Container registry
kubeflow-	Machine Learning service workspace
kubeflowa	Storage account
kubeflowa	Key vault
kubeflowa	Application Insights

The Code

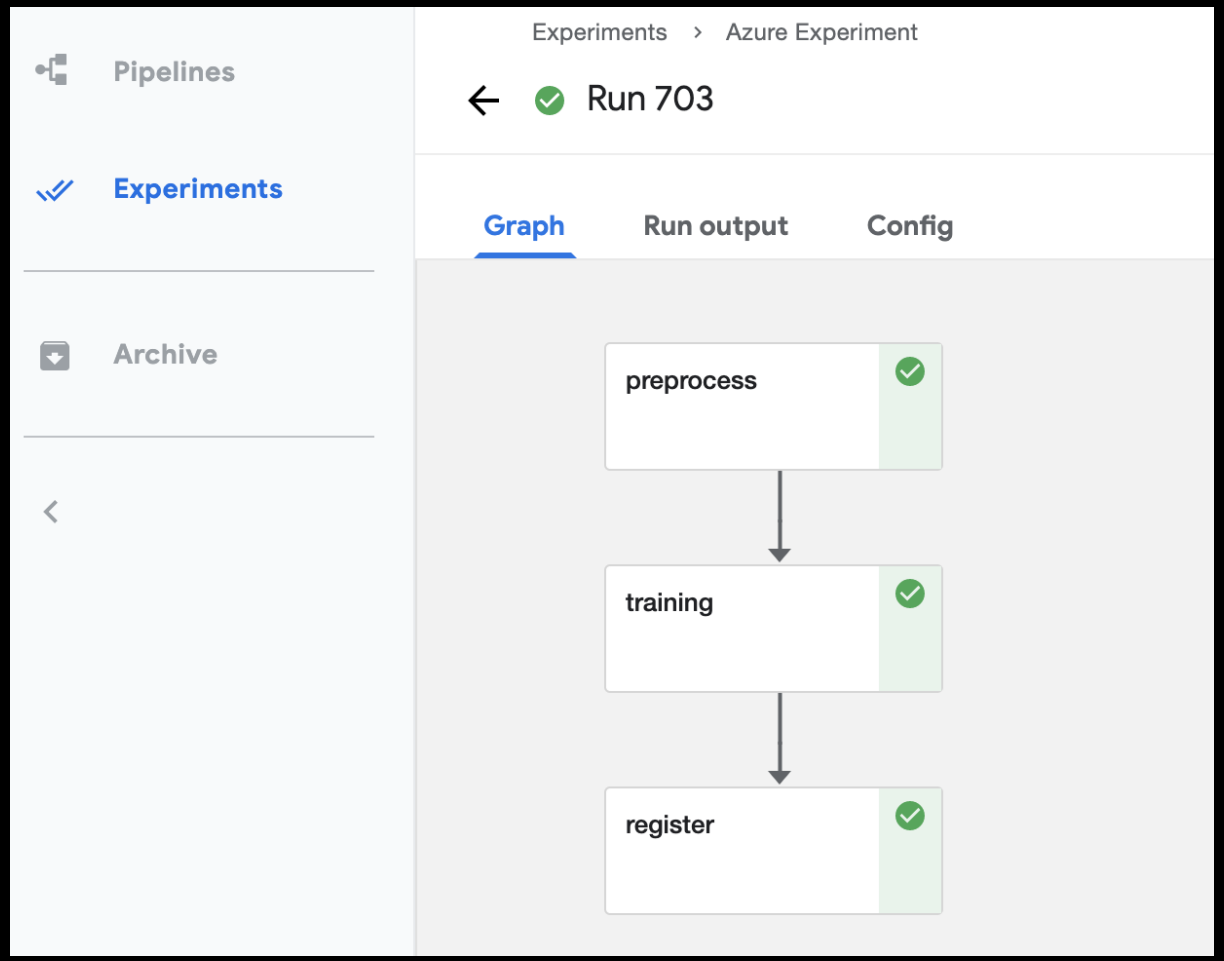

EXPLORER
data.py
OPEN EDITORS
data.py code\preprocess 4
KUBEFLOW-AND-MLOPS
code
deploy
preprocess
build.sh
data.py 4
Dockerfile
requirements.txt
register
training
pipeline.py
pipeline.yaml
kubernetes
model
readme-images
release-pipelines
.gitignore
azure-pipelines-tnb.yml
azure-pipelines.yml
LICENSE
code > preprocess > data.py > download
import os
import time
import shutil
import wget
import zipfile
import argparse
import numpy as np
import tensorflow as tf
from pathlib2 import Path
def check_dir(path, check=False):
if check:
assert os.path.exists(path), '{} does not exist!'.format(path)
else:
if not os.path.exists(path):
os.makedirs(path)
return Path(path).resolve(strict=False)
def download(source, target, force_clear=False):
if force_clear and os.path.exists(target):
print('Removing {}...'.format(target))
shutil.rmtree(target)
check_dir(target)

Build Pipeline

https://github.com/DivineOps/kubeflow-and-mlops

```yaml
stages:
- stage: ContainerConfig
  displayName: 'Configure and Register Containers'
  jobs:
  - job: Containers
    pool:
      name: Hosted Ubuntu 1604
    steps:
    - task: AzureCLI@1
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          az acr login -n $(KF_ACR)
          cd code
          cd preprocess
          docker build -t $(KF_ACR).azurecr.io/preprocess:$BUILD_SOURCEVERSION .
          docker push $(KF_ACR).azurecr.io/preprocess:$BUILD_SOURCEVERSION
        displayName: 'Build & Push Preprocess Image'

    - task: AzureCLI@1
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          cd code
          cd training
          docker build -t $(KF_ACR).azurecr.io/training:$BUILD_SOURCEVERSION .
          docker push $(KF_ACR).azurecr.io/training:$BUILD_SOURCEVERSION
        displayName: 'Build & Push Training Image'

    - task: AzureCLI@1
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          cd code
          cd register
          docker build -t $(KF_ACR).azurecr.io/register:$BUILD_SOURCEVERSION .
          docker push $(KF_ACR).azurecr.io/register:$BUILD_SOURCEVERSION
        displayName: 'Build & Push Register Image'
```

```yaml
- stage: KubeflowTrigger

  variables:
    KF_BATCH: 32
    KF_EPOCHS: 5
    KF_LEARNING_RATE: 0.0001
    KF_MODEL_NAME: tacosandburritos
    KF_PERSISTENT_VOLUME_NAME: azure
    KF_PERSISTENT_VOLUME_PATH: /mnt/azure

  dependsOn: ContainerConfig
  displayName: 'Trigger Kubeflow Pipeline'
  jobs:
  - job: Kubeflow
    pool:
      name: Hosted Ubuntu 1604
    steps:
    - task: AzureCLI@1
      env:
        KF_MAPPED_SERVICE_PRINCIPAL_PASSWORD: $(KF_SERVICE_PRINCIPAL_PASSWORD)
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          az aks get-credentials -g $(KF_RESOURCE_GROUP) -n $(KF_AKS_CLUSTER)
          kubectl port-forward --namespace kubeflow svc/ml-pipeline 8888:8888 &
          kubepid=$!

          sudo apt-get install python3-setuptools
          pip3 install wheel
          pip3 install kfp

          touch script.py
          echo "import kfp" >> script.py
          echo "client = kfp.Client(host='localhost:8888')" >> script.py
          echo "client.run_pipeline('$KF_EXPERIMENT_ID', 'Run ${BUILD_BUILDID}', params={'imagetag': '${BUILD_SOURCEVERSION}',

          cat script.py

          python3 script.py

          kill $kubepid
        displayName: 'Trigger Kubeflow Pipeline'
```

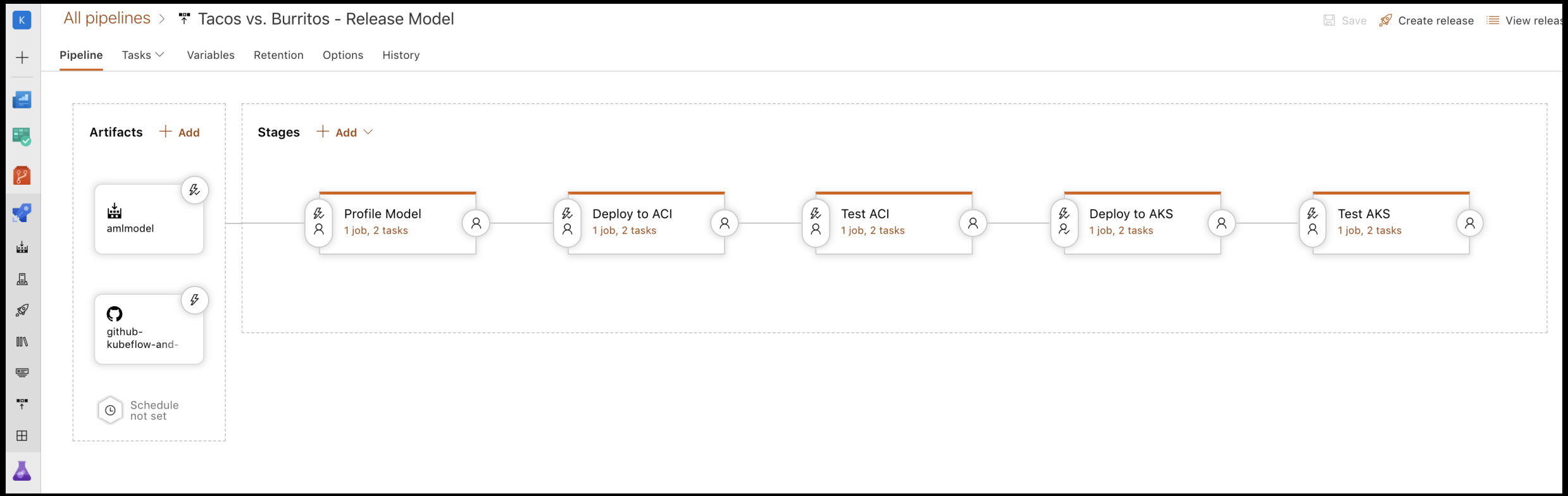


Kubeflow Pipeline

https://www.kubeflow.org/docs/azure/azureendtoend/

Release Pipeline

https://dev.azure.com/sasrose/kubeflow/_release?_a=releases&view=mine&definitionId=3



Even a simple CI/CD pipeline is better than none!

AI Ethics

Bias is a property of information

Build AI responsibly!

Thank You!

@DivineOps

Resources

GitHub repo

https://www.kubeflow.org/docs/azure/azureendtoend/

Deploy Kubeflow on Azure

https://www.kubeflow.org/docs/azure/deploy/install-kubeflow/

Example Kubeflow Azure Pipeline

https://www.kubeflow.org/docs/azure/azureendtoend/

Release pipeline

https://dev.azure.com/sasrose/kubeflow/_release