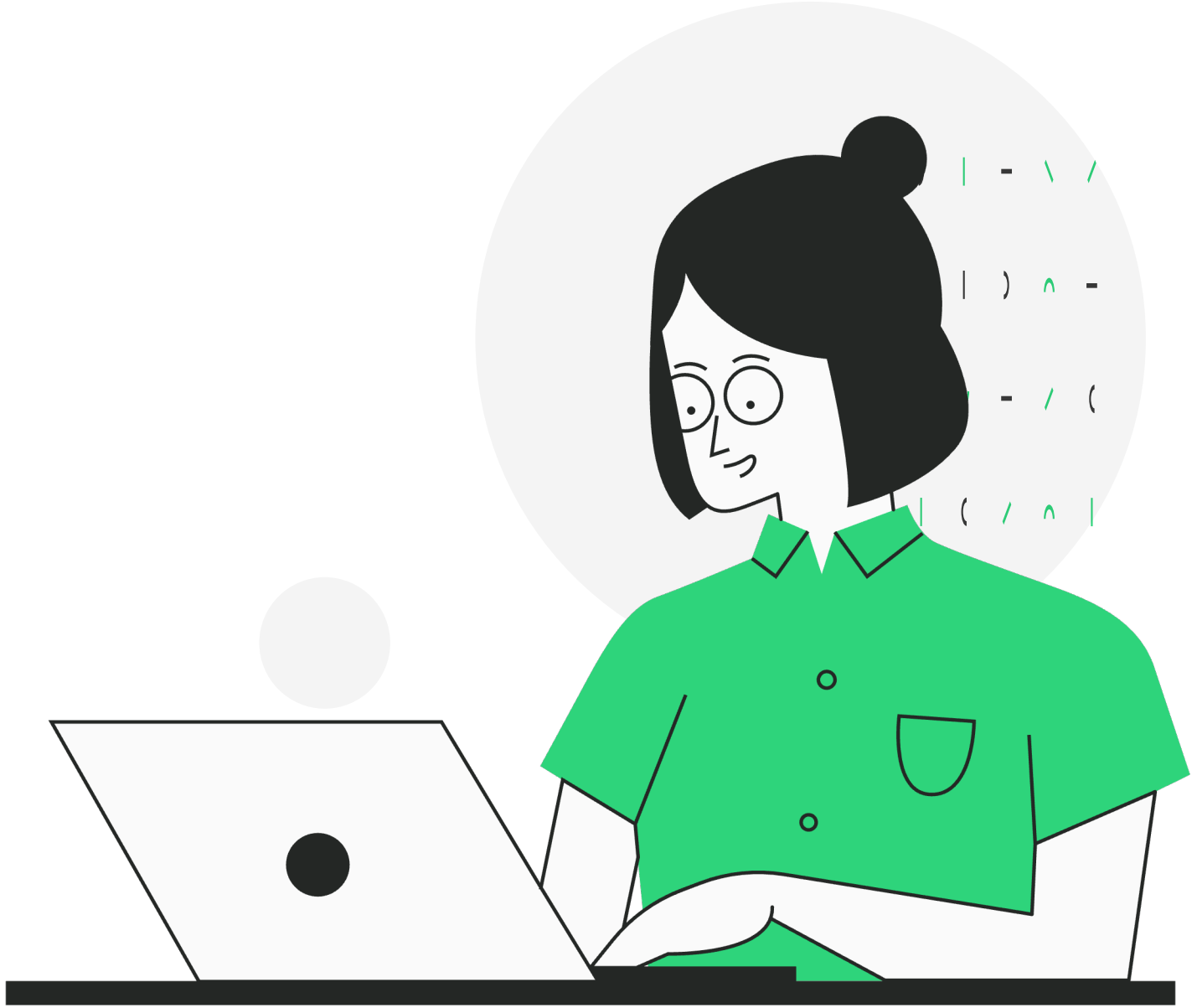
Introduction to MLOps



The MLOps Project

Author Introduction

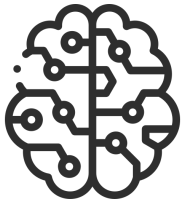
Lavi Nigam

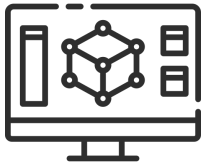
Data Scientist @ Gartner

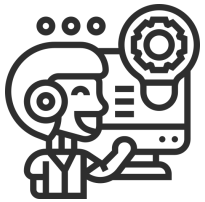
The MLOps Project

What are we going to learn today?

What we Have?	The New Stuff?	What we should have?
• Current ML Pipeline • New Age ML Pipeline	• Model & Data Versioning • Model Interpretability • AutoML • Multi Model Analysis • Model & Data Bias • Distributed Data Science • Data Interactive Apps	• MLOps • Continuous Deployment & Integration • Model Production API

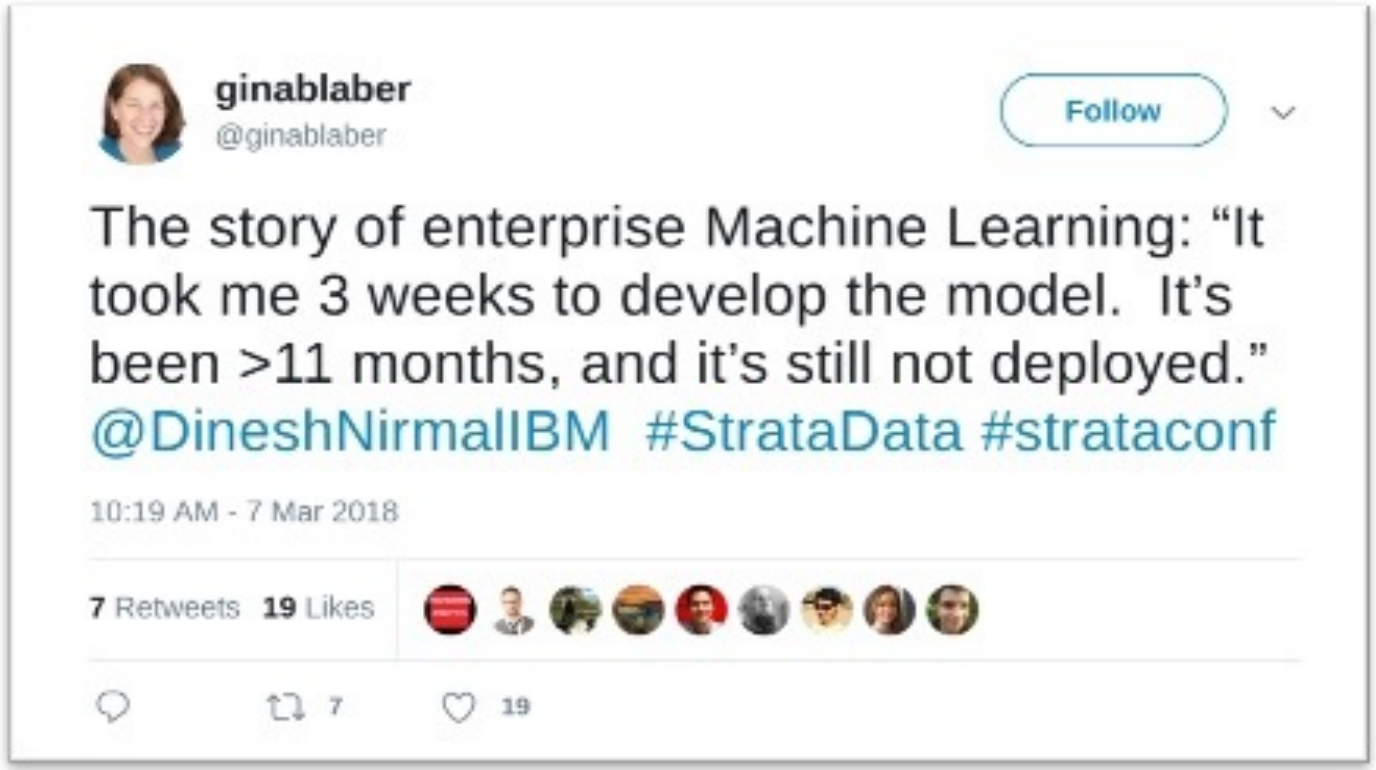






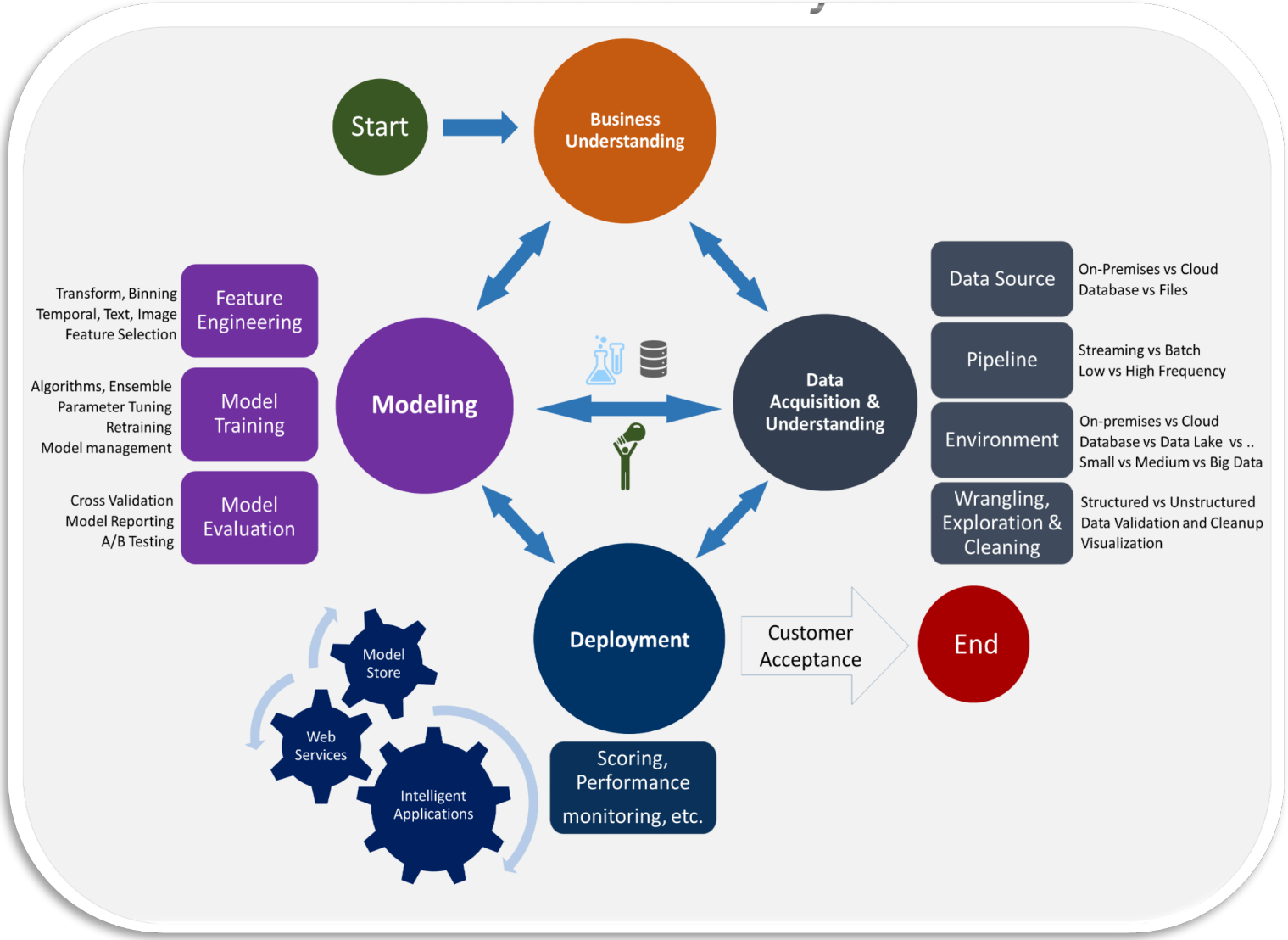
The MLOps Project

Alright, but why do we care for all of this?



But, isn't the Data Science we do already Enterprise Ready?

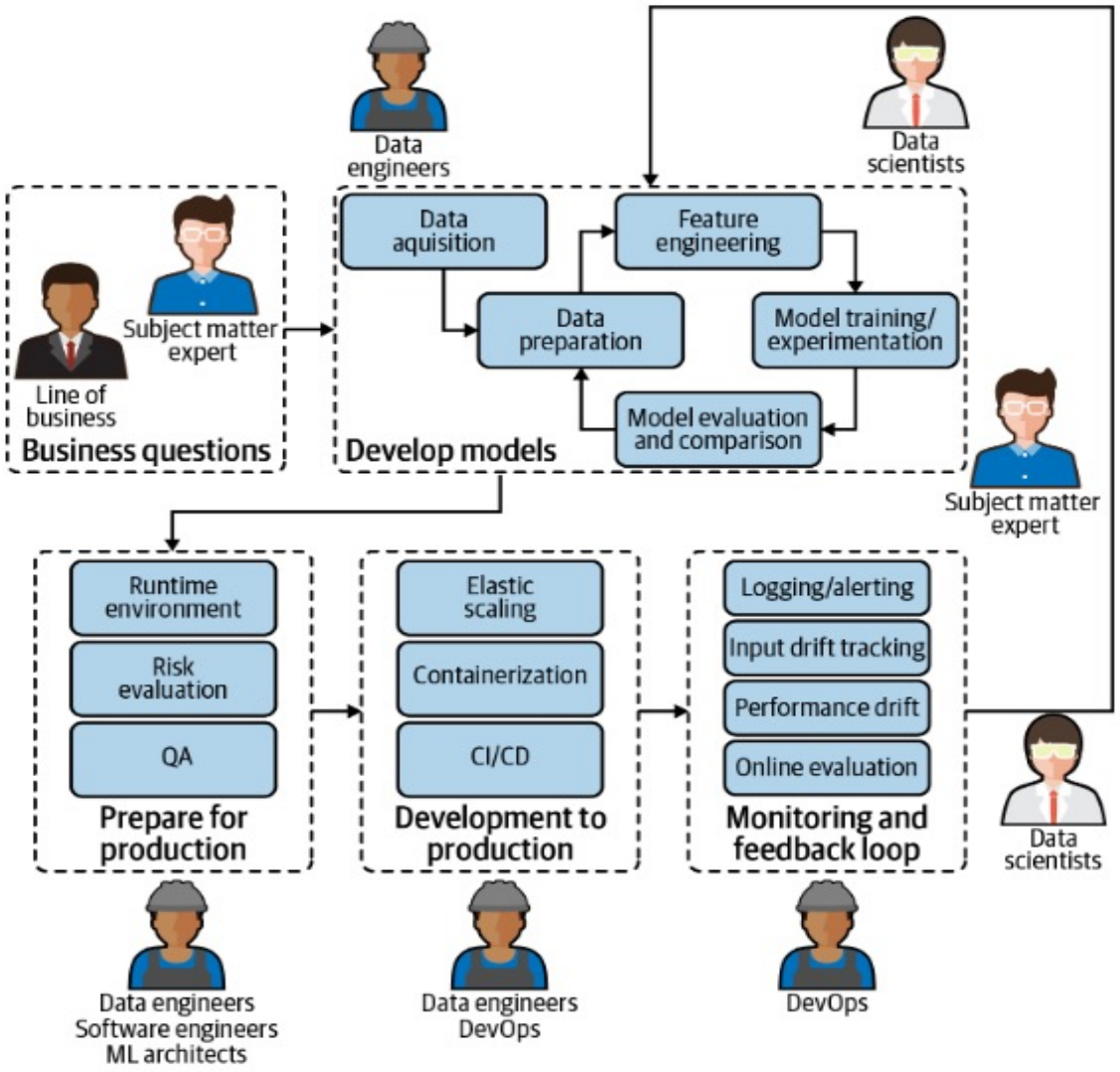
The Current ML Pipeline



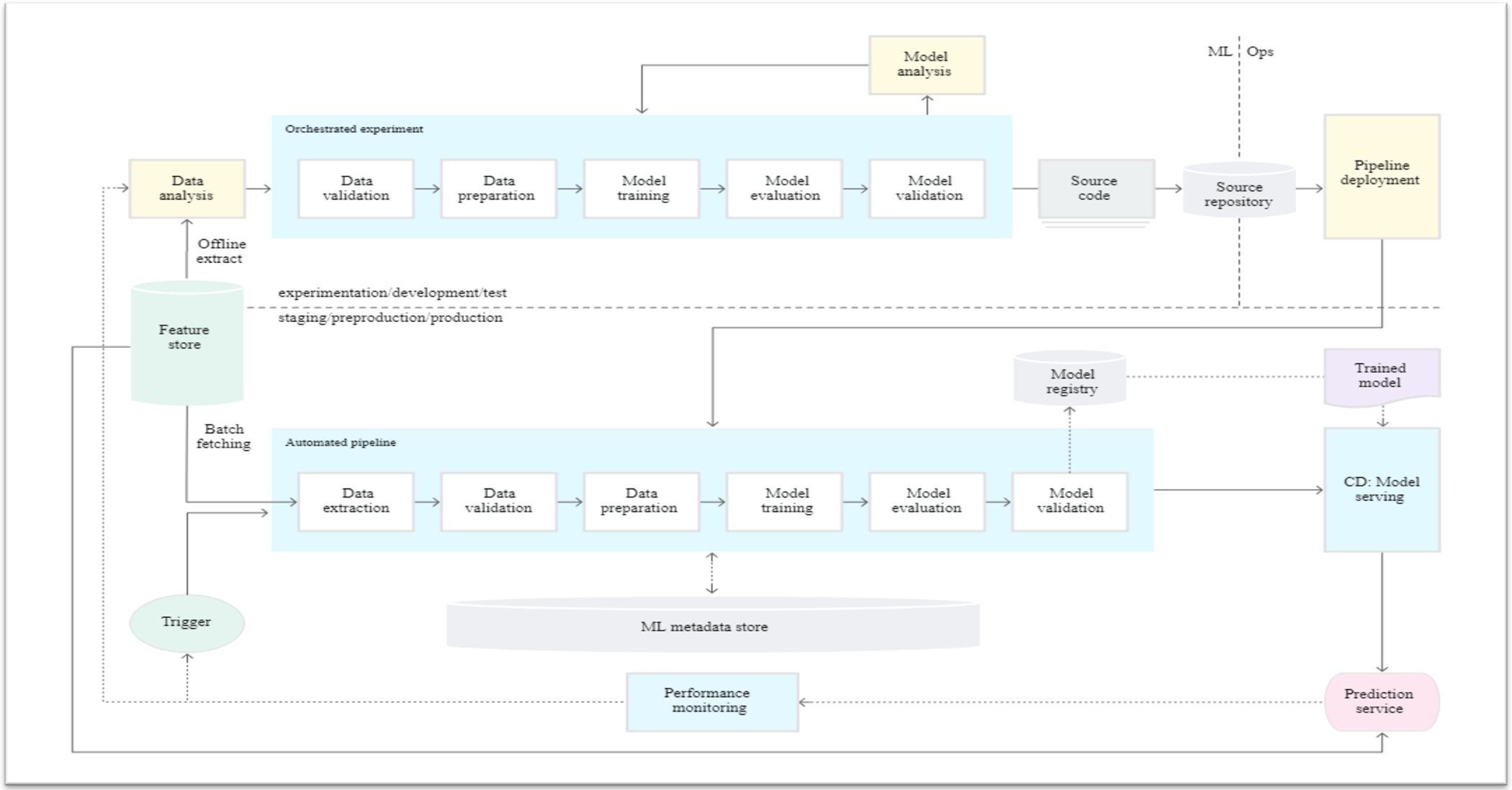
https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/media/lifecycle/tdsp-lifecycle2.png

The MLOps Project

New Age ML Pipeline



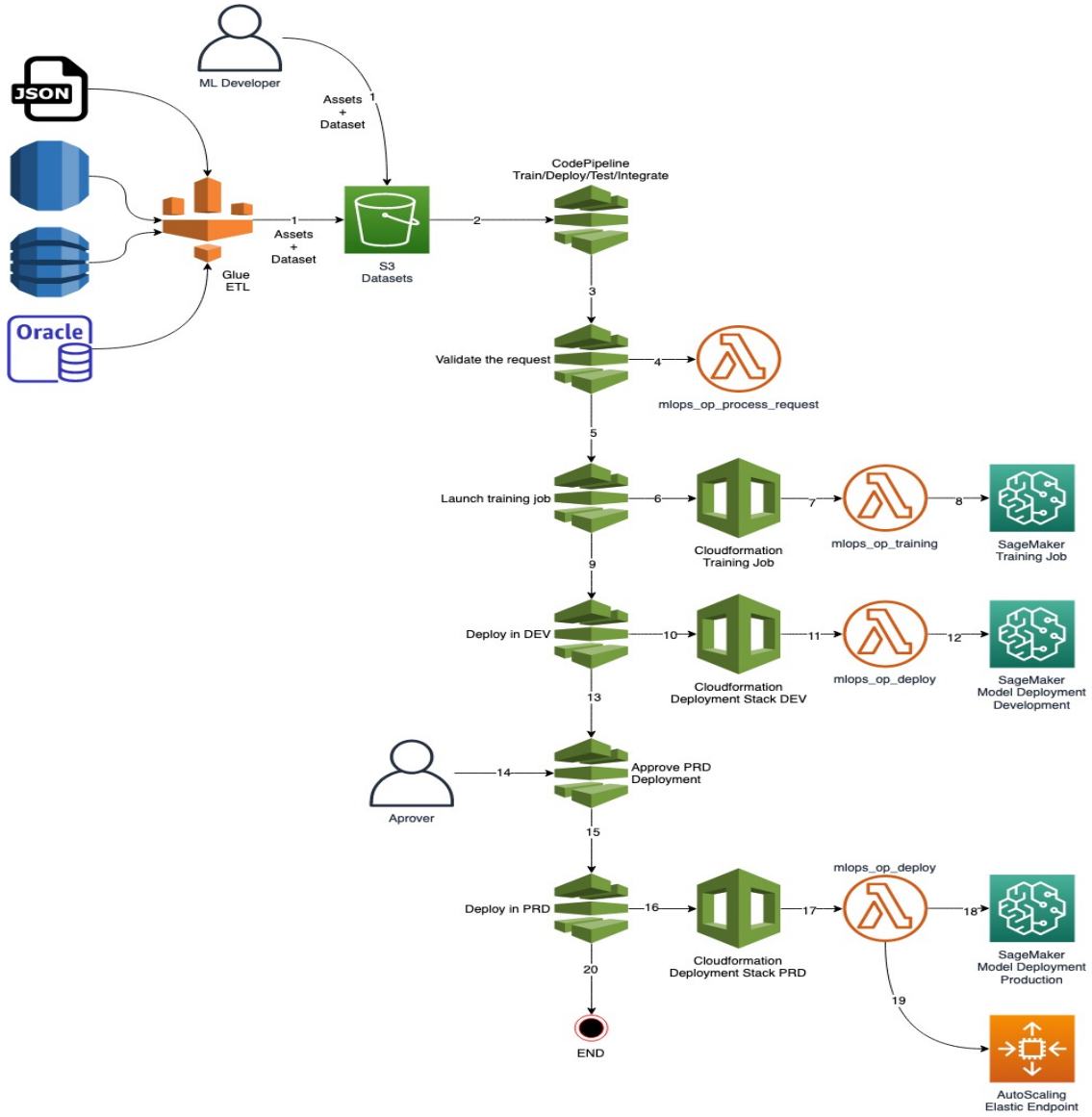
New Age ML Pipeline - GCP



https://cloud.google.com/solutions/images/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning-3-ml-automation-ct.svg

The MLOps Project

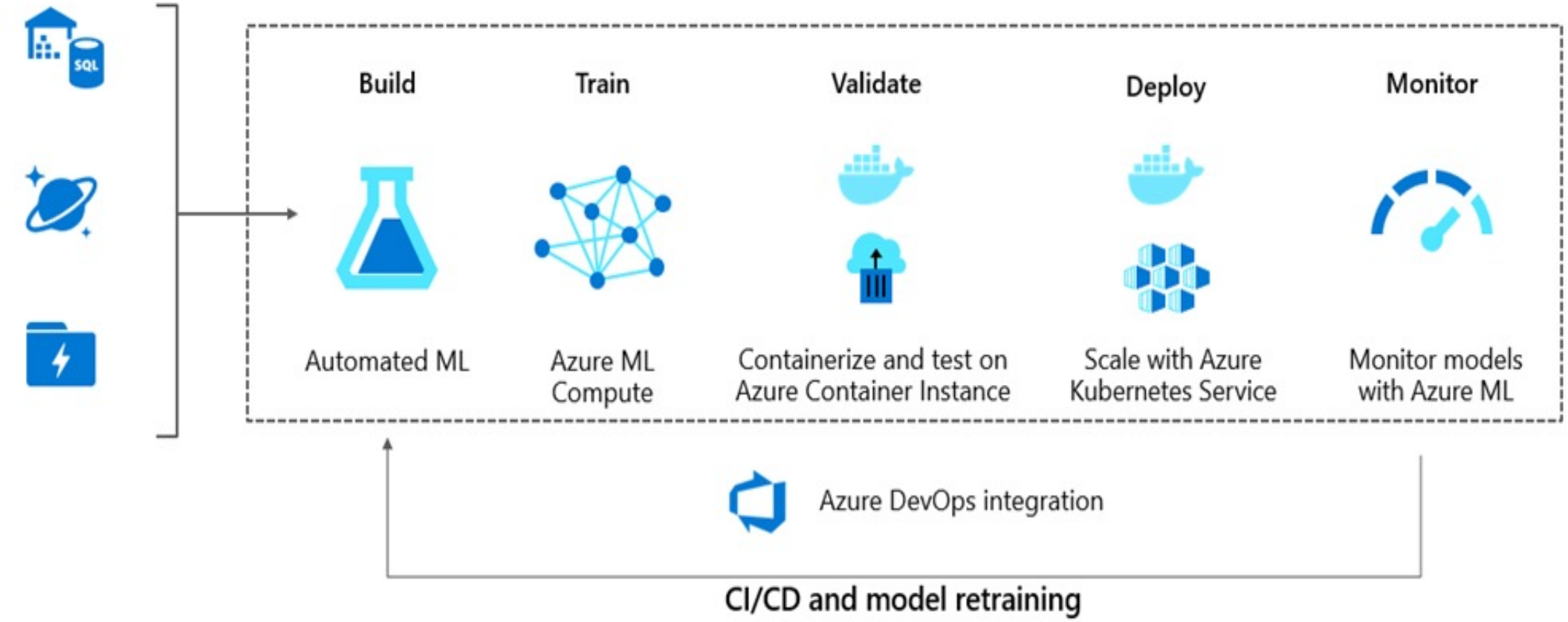
New Age ML Pipeline - AWS



https://raw.githubusercontent.com/awslabs/amazon-sagemaker-mlops-workshop/master/imgs/MLOps_Train_Deploy_TestModel.jpg

The MLOps Project

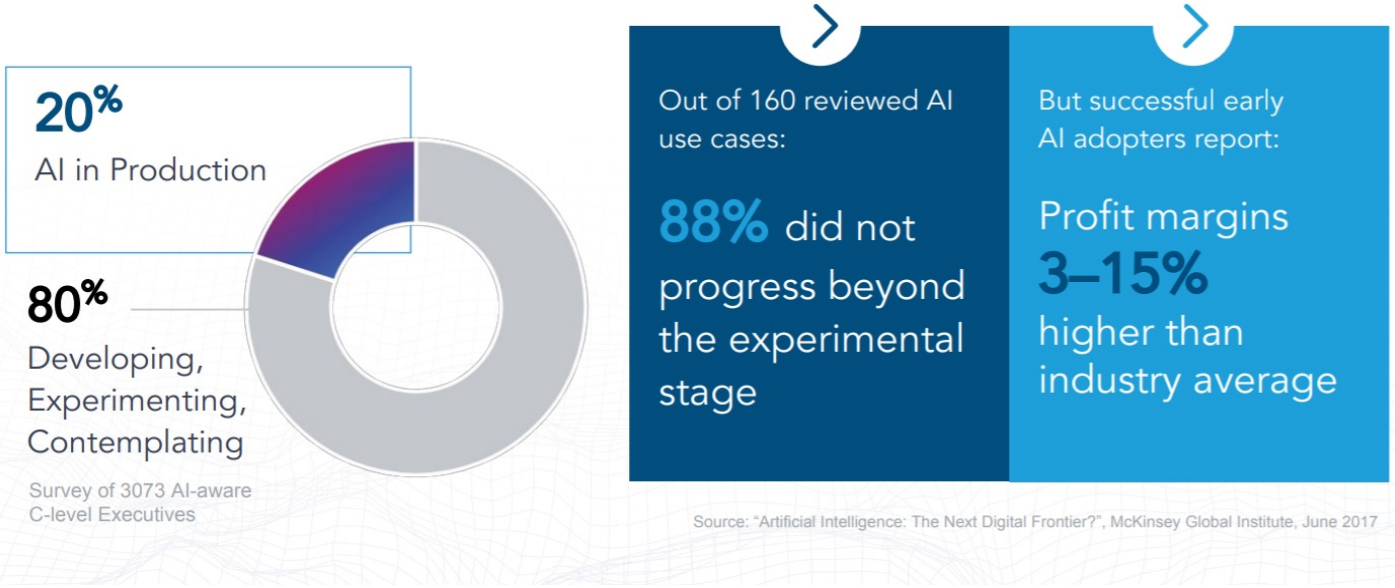
New Age ML Pipeline - Azure



https://azurecomcdn.azureedge.net/mediahandler/acomblog/media/Default/blog/575cb9f3-2beb-4c90-a3b7-a8ee645ce24e.png

Alright, so many new things! Who says we need them?

What the experts say



"As the market for artificial intelligence (AI) technologies and techniques matures and grows, users and partners require more and better access to innovative AI models, applications and platforms."

- ***Gartner***, *Emerging Technologies and Trends Impact Radar: Artificial Intelligence*, 13 November 2019

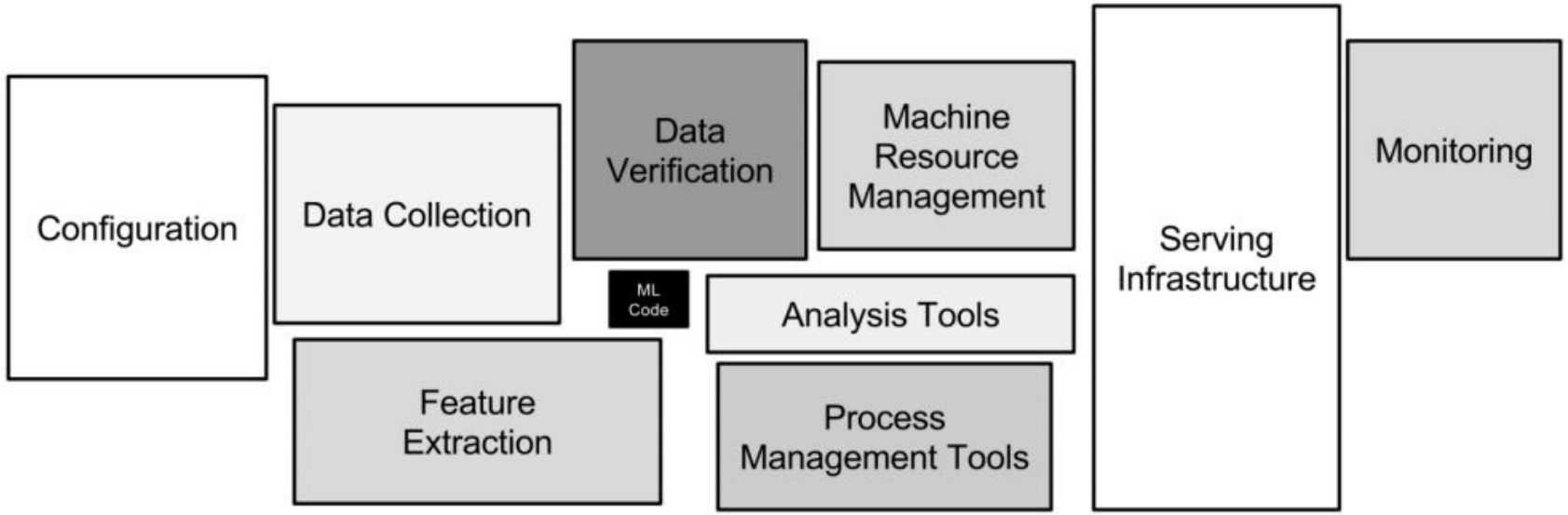
"Technology innovation leaders are keen to apply DevOps principles for AI and ML projects, but they often struggle with architecting a solution for automating end-to-end ML pipelines across data preparation, model building, deployment and production due to lack of process and tooling know-how."

- **Gartner**, *Accelerate Your Machine Learning and Artificial Intelligence Journey Using These DevOps Best Practices*, 12 November 2019, Arun Chandrasekaran and Farhan Choudhary

The Bigger Picture

Hidden Technical Debt in Machine Learning Systems

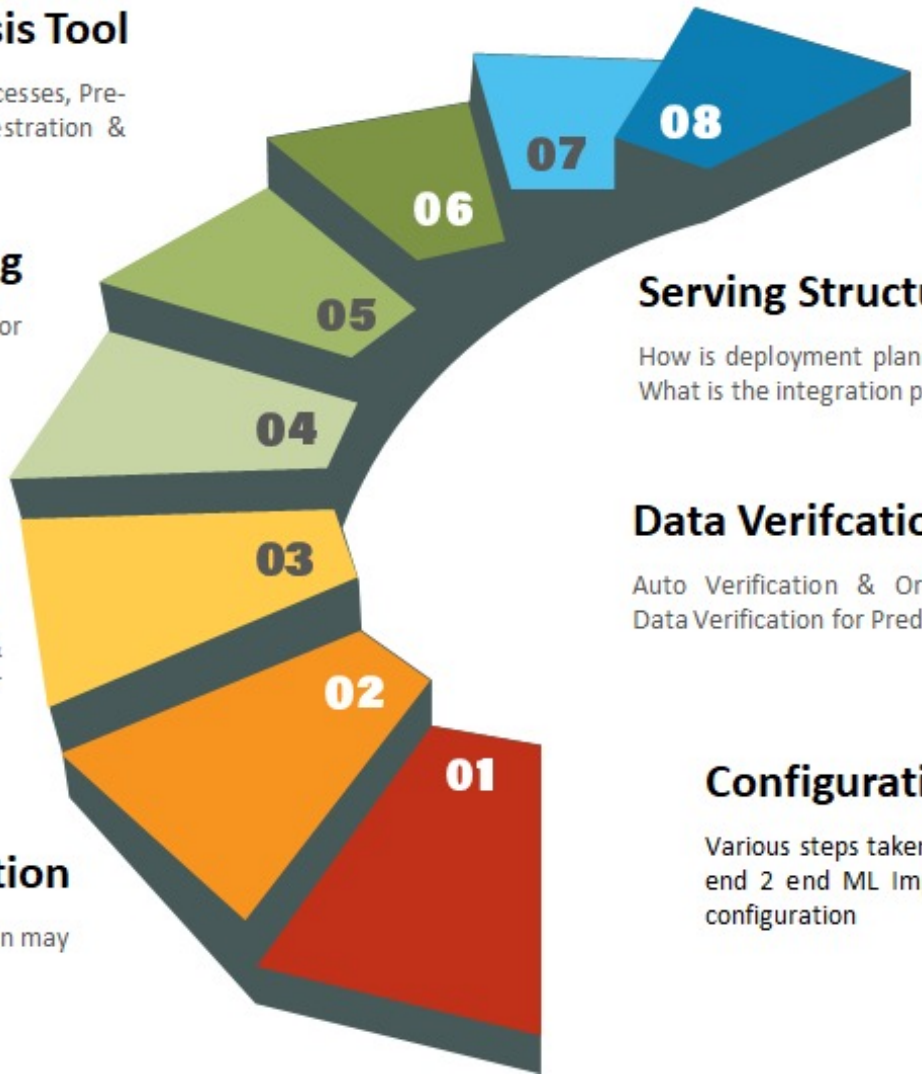
D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.



https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

Fine, I am sold! Take my money & explain these - **PLEASE!**

Real World Machine Learning



Process Management & Analysis Tool
08

Feature Engineering, Pre-Processes, Pre-Model & Post-Model Orchestration & Automation

Feature Extraction
07

Do we have feature store that tracks all feature extraction across experiments?

06 Monitoring

How do you monitor your models for relevance and error for new data?

Serving Structure
05

How is deployment planned for model? What is the integration plan?

04 Machine Resource Mangement

Compute Infrastrcure for Training & Prediction. Server Requirement for Deployement

Data Verifcation
03

Auto Verification & Orchestration of Data Verification for Predictions

02 Data Collection

Source and type of data collection may change over time

Configuration
01

Various steps taken during each step of end 2 end ML Implementation. Lot of configuration

The MLOps Project

What is MLOPs

Extension to DevOps

MLOPS is extension of DevOps, software codes don't degrade ML models does.

ML & DS asset

Your python code, model, features , hyperparameters, charts can be saved for later purpose

Practices , Frameworks & Tools

MLOps is a combination of multiple practices and frameworks to facilitate continuous monitoring and retraining so that the algorithms may continue to be used in production

Maintainability, Testability & Scalability

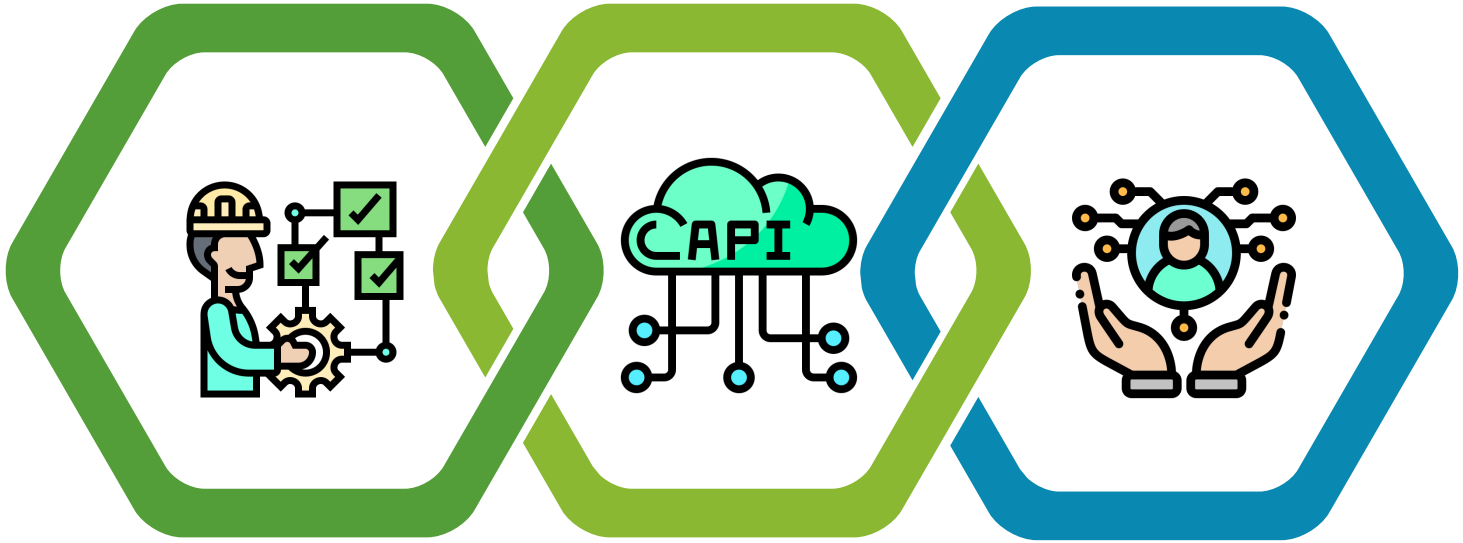
AI models should be scalable, its should be tested and maintained by properly monitoring the overall process

Ethics, Bias & Fairness

Design algorithms that make fair predictions, it should remove data and model bias. AI system make important decisions, we need to be aware of limitations

The MLOps Project

What MLOps is Not



Jupyter Notebooks in production environment

Building API of your Model

Considering models as First Class citizen

Features of MLOps

Model Development

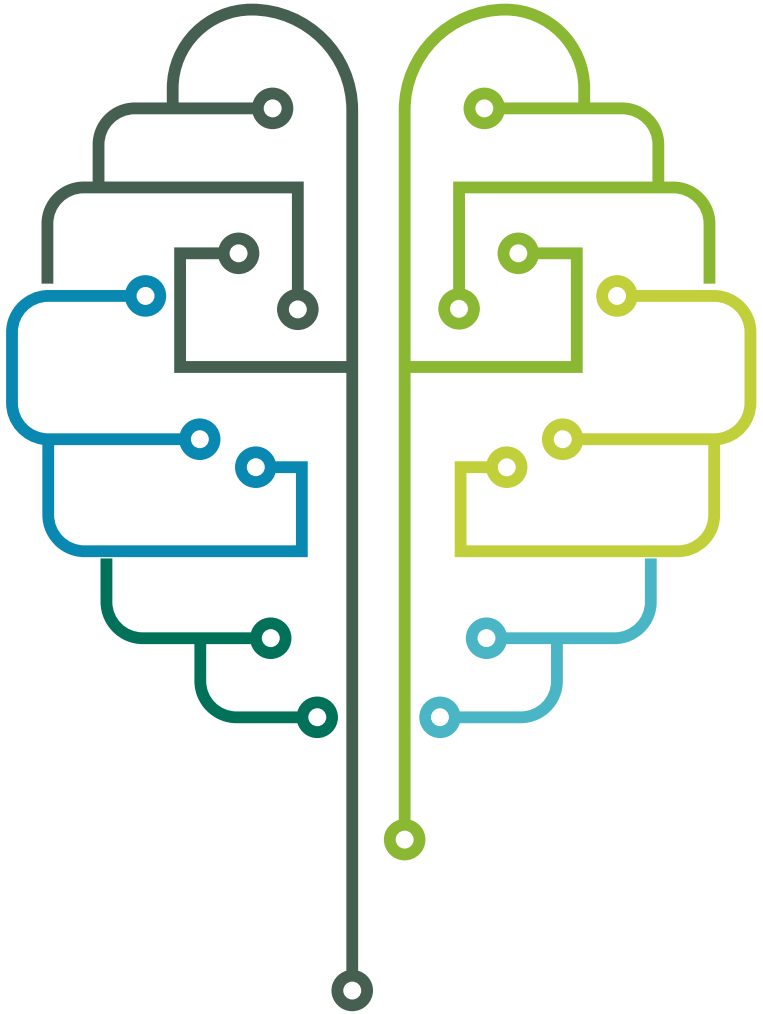
Model development is an iterative process, in which many models are derived, tested and built upon until a model fitting the desired criteria is built

Production & Deployment

Model Deployment involves making the model available in production environments, so they can be used to make predictions and provide value to other systems

Iteration and Life cycle

Breaks ML process into smaller steps like Data Collection, Data Normalization, Data Modelling, Model Training and Feature Engineering & Deploying Models to Production



Governance, Continuous Integration & Continuous learning

Set of activities, policies and procedures which formalize model and model risk management activities. MLOPS uses CI/CD practices that enable application development teams to deliver code changes more frequently and reliably.

Model Interpretation

Explainable AI is a set of tools and frameworks to help you understand and interpret predictions made by your machine learning models

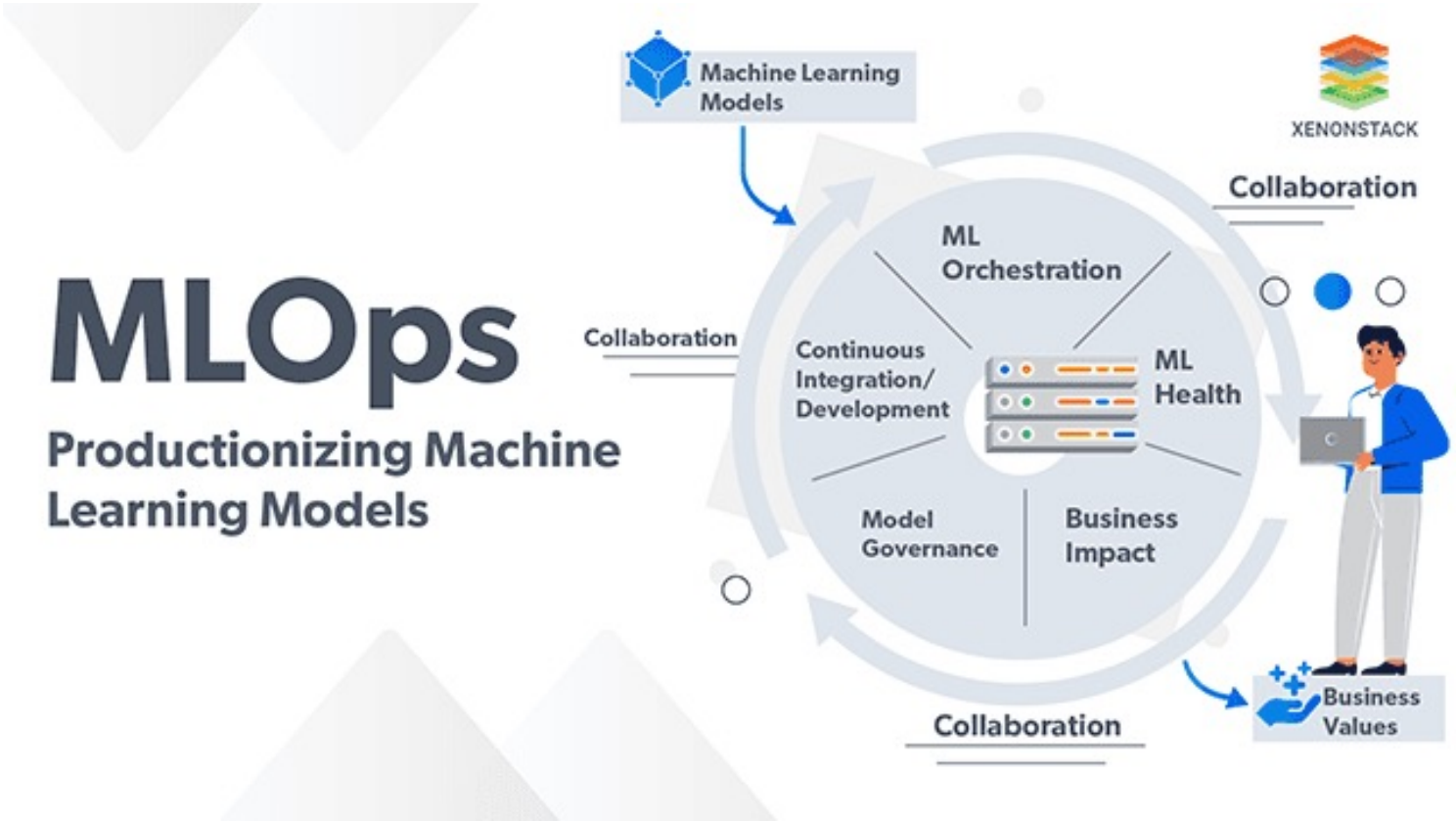
Model and Data Monitoring

Monitoring should be designed to provide early warnings to the myriad of things that can go wrong with a production ML model

ML Unique Challenge in Production

Dataset dependency

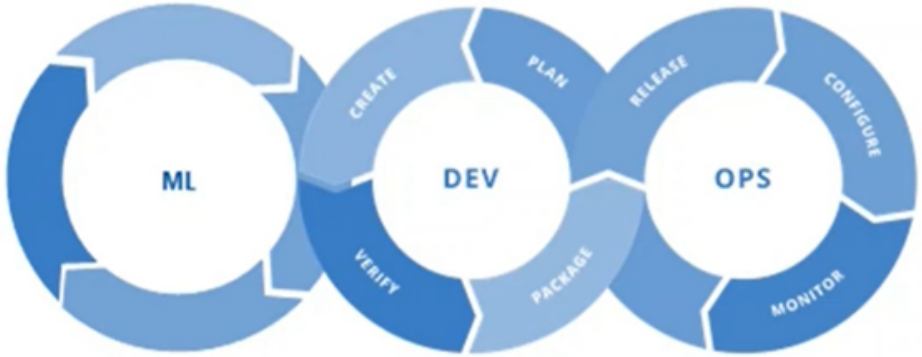
- ML ‘black box’ into which many inputs (algorithmic, human, dataset etc.) go to provide output.
- Difficult to have reproducible, deterministically ‘correct’ result as input data changes
- ML in production may behave differently than in developer sandbox because live data ≠ training data



The MLOps Project

ML Unique Challenge in Production

MLOps = ML + DEV + OPS



Experiment
Data Acquisition
Business Understanding
Initial Modeling

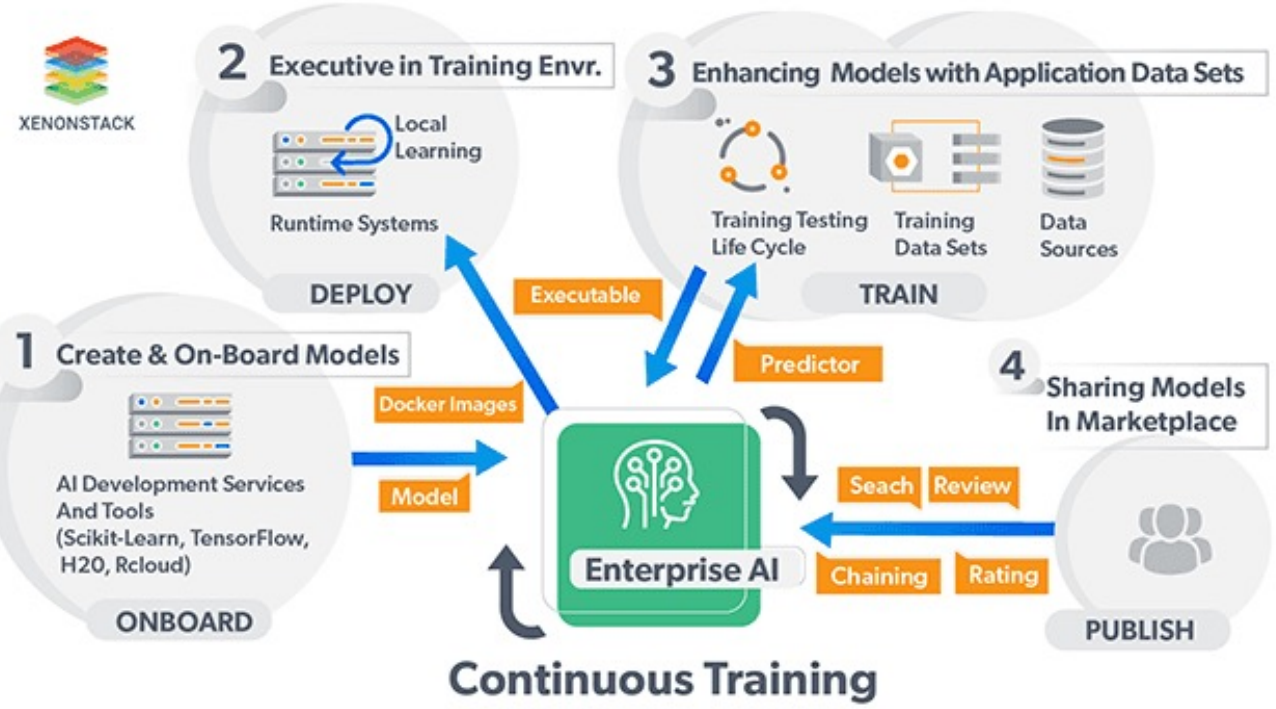
Develop
Modeling + Testing
Continuous Integration
Continuous Deployment

Operate
Continuous Delive
Data Feedback Lo
System + Model M

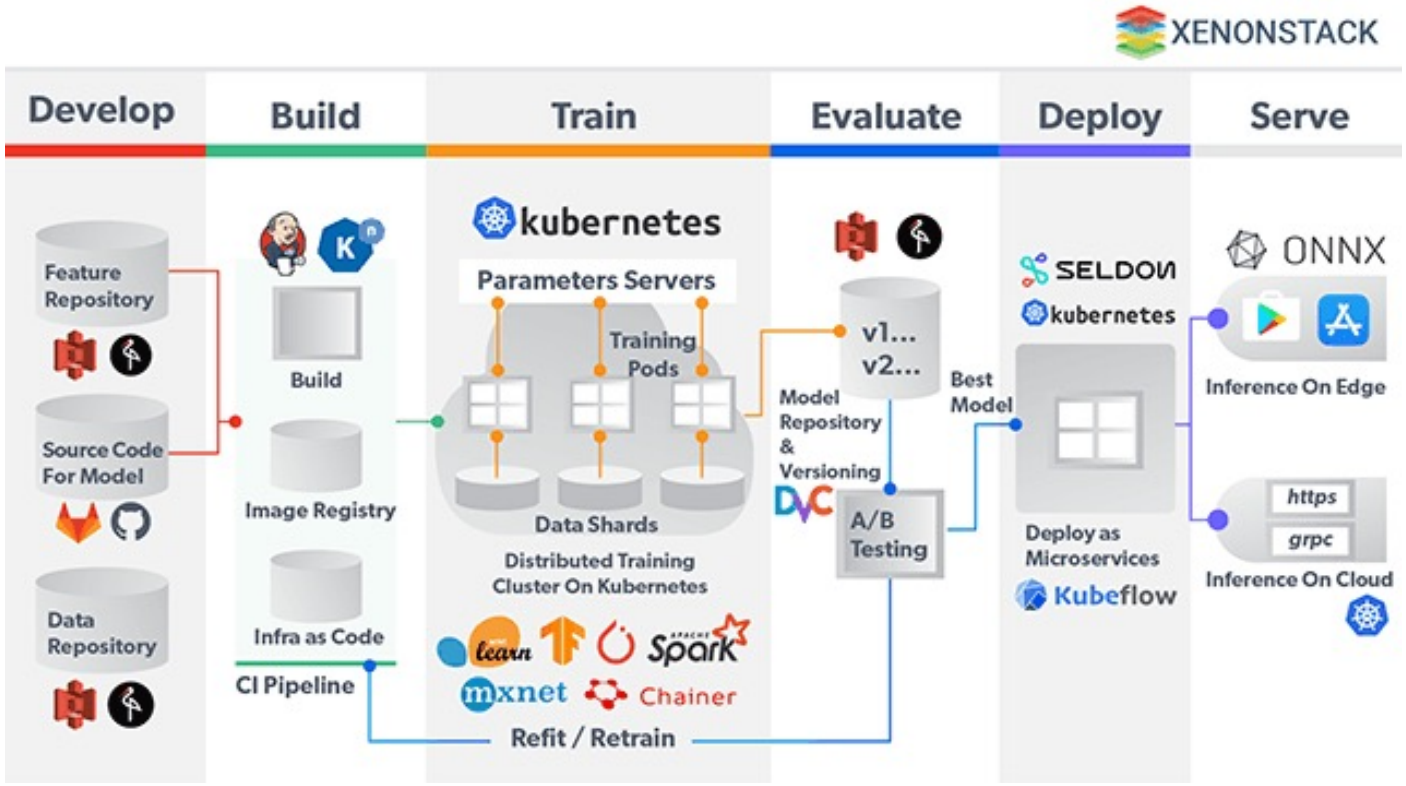
- Multiple loosely coupled pipelines running possibly in parallel, with dependencies and human interactions
- Feature engineering pipelines must match for Training and Inference
- Further complexity if ensembles, federated learning etc are used

ML Unique Challenge in Production

- Possibly differing engines (Spark, TensorFlow, Caffe, PyTorch, Sci-kit Learn, etc.)
- Different languages (Python, Java, Scala, R ..)
- Inference vs Training engines
- Inference (Prediction, Model Serving) can be REST endpoint/custom code, streaming engine, micro-batch, etc.
- Feature manipulation done at training needs to be replicated (or factored in) at inference
- Each engine presents its own scale opportunities/issues



ML Unique Challenge in Production



- Model Risk Management
- Reproducing and Explaining ML Decisions
- Algorithm Fairness Monitoring

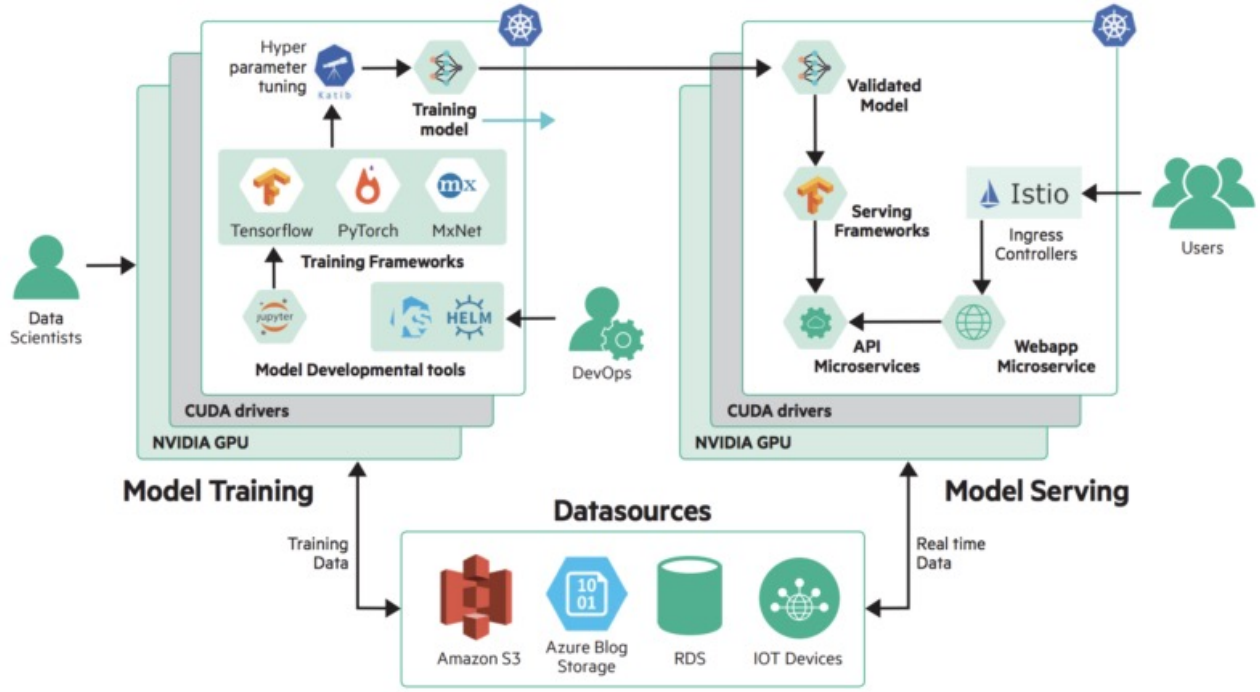
ML Unique Challenge in Production

COLLABORATION

- Expertise mismatch between Data Science & Ops complicates handoff and continuous management and optimization

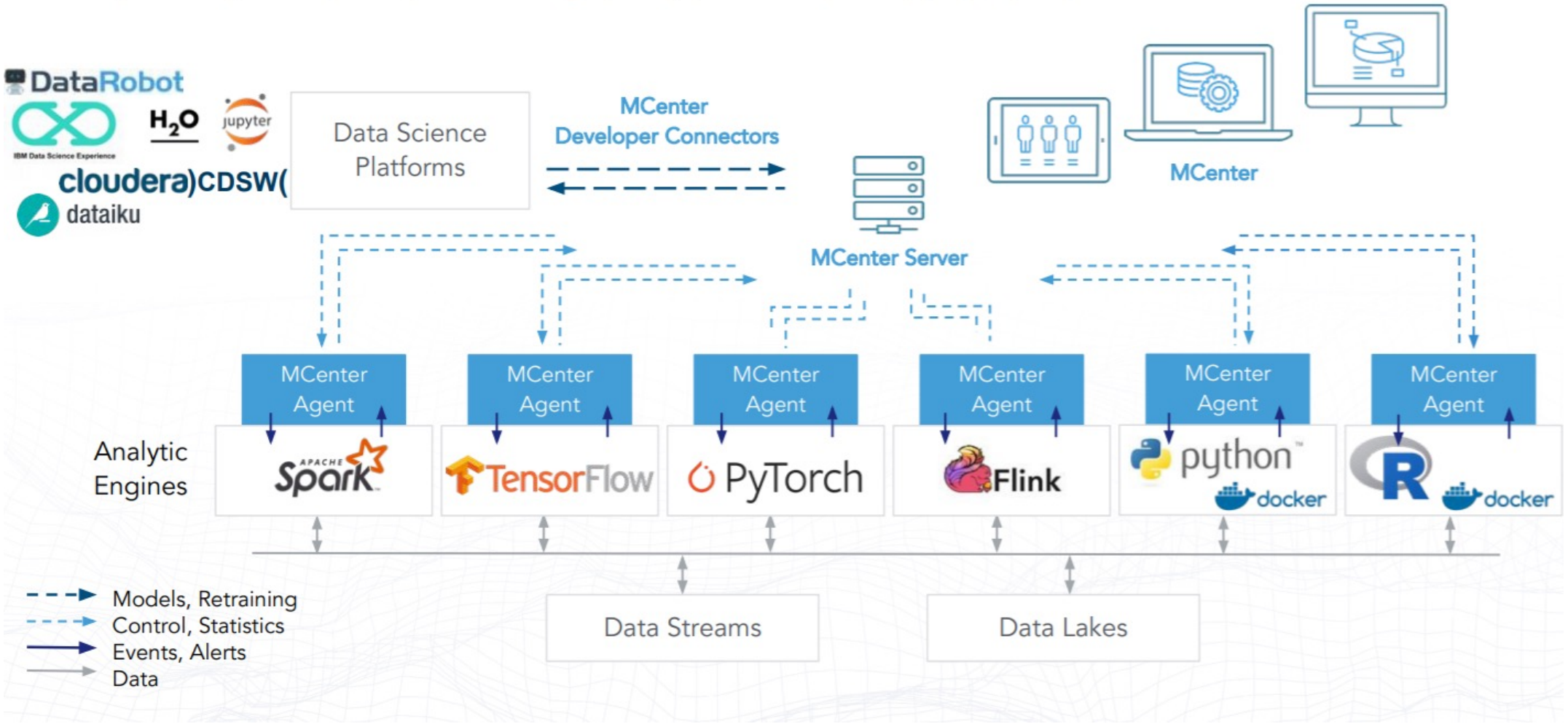
PROCESS

- Many objects to be tracked and managed (algorithms, models, pipelines, versions etc.)
- ML pipelines are code. Some approach them as code, some not
- Some ML objects (like Models and Human approvals) are not best handled in source control repositories



The MLOps Project

Final MLOps Architecture



Production ML
Avengers Team

Data/ML Pipelines

Data Versioning

Experiment Management

Model Management

Model Monitoring

The MLOps Project

Model & Data Versioning

What is it?

A version control method to version the large files such as datasets and trained model files

How to do it?

You can use DVC Tool that helps you manage/track your code and data together.

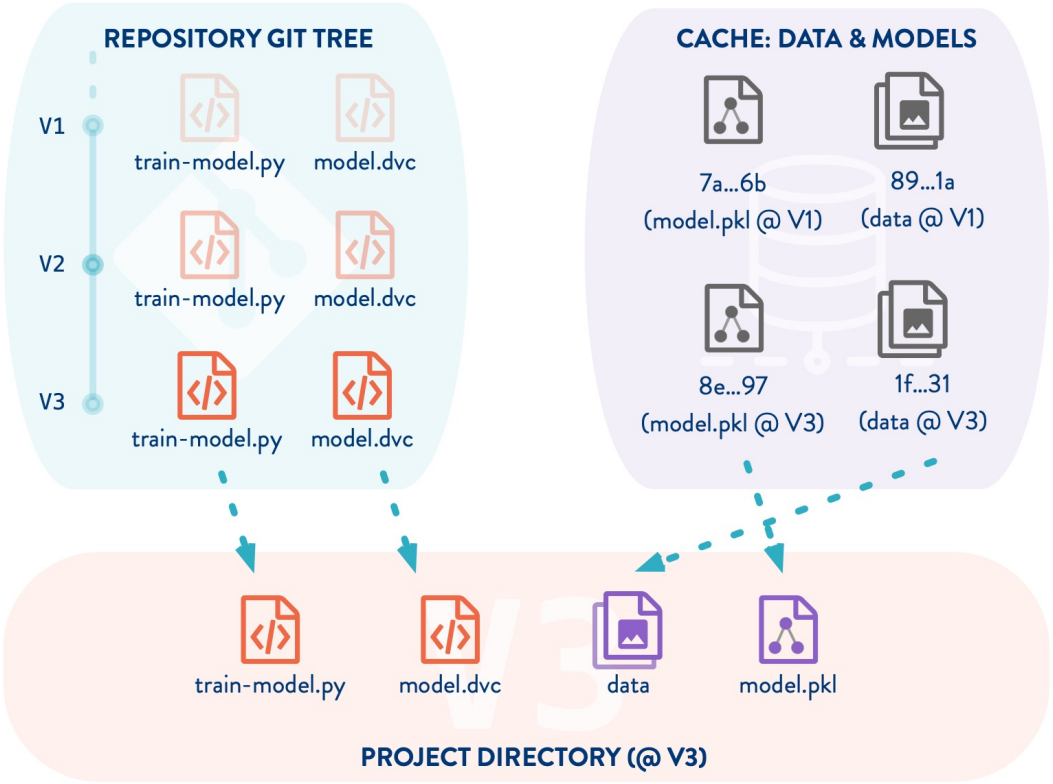
Why do it?

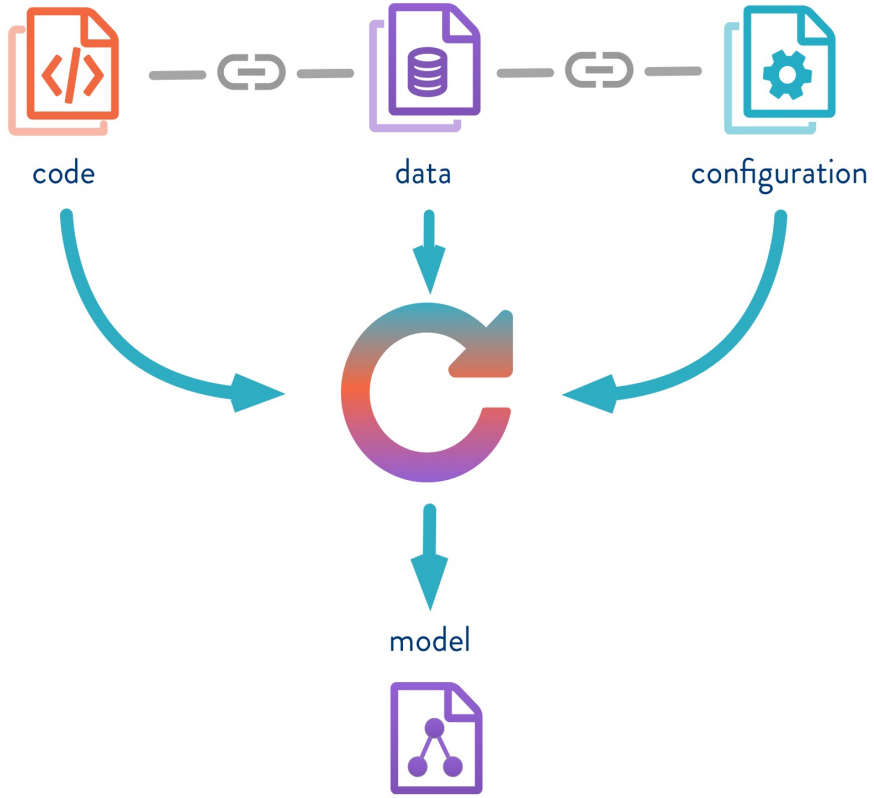
It helps with the reproducibility of artifacts of different ML experiments

Give me the link?

https://dvc.org/

Production in an enterprise





The MLOps Project

Model Interpretation

What is it?

Model Interpretability means explaining decisions made my black-box models

How to do it?

SHAP,EII5,LIME

Why do it?

Optimize for true objectives, understand unexpected behavior and debug model for better decisions

Give me the link?

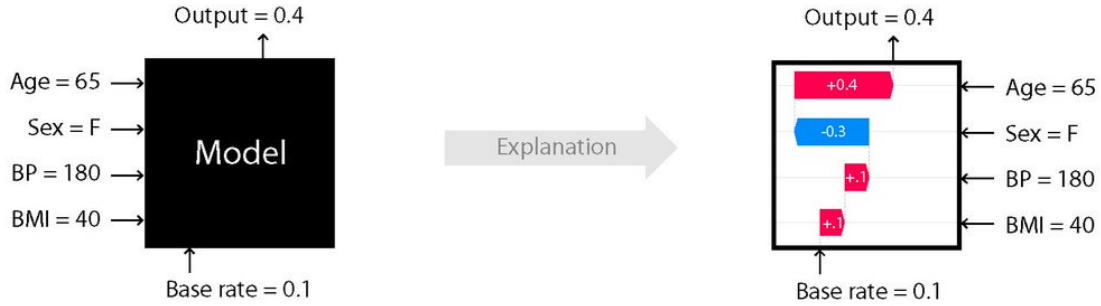
https://eli5.readthedocs.io/en/latest/

https://shap.readthedocs.io/en/latest/

Model Interpretation using SHAP

Weight	Feature
0.0750 ± 0.1159	Goal Scored
0.0625 ± 0.0791	Corners
0.0437 ± 0.0500	Distance Covered (Kms)
0.0375 ± 0.0729	On-Target
0.0375 ± 0.0468	Free Kicks
0.0187 ± 0.0306	Blocked
0.0125 ± 0.0750	Pass Accuracy %
0.0125 ± 0.0500	Yellow Card
0.0063 ± 0.0468	Saves
0.0063 ± 0.0250	Offsides
0.0063 ± 0.1741	Off-Target
0.0000 ± 0.1046	Passes
0 ± 0.0000	Red
0 ± 0.0000	Yellow & Red
0 ± 0.0000	Goals in PSO
-0.0312 ± 0.0884	Fouls Committed
-0.0375 ± 0.0919	Attempts
-0.0500 ± 0.0500	Ball Possession %

SHAP



The MLOps Project

Automated Machine Learning

What is it?

Process of Automating end-to-end ML workflow

Why do it?

Helps in faster feature creation and model building

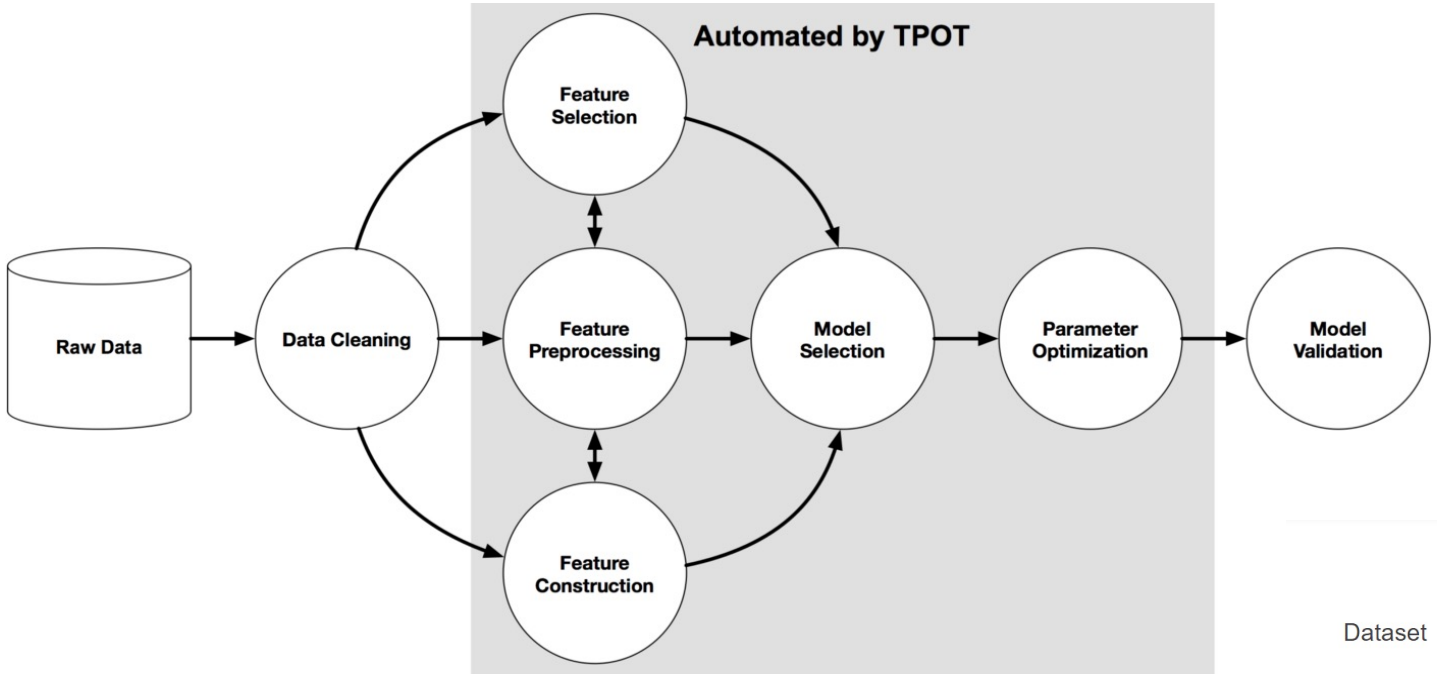
How to do it?

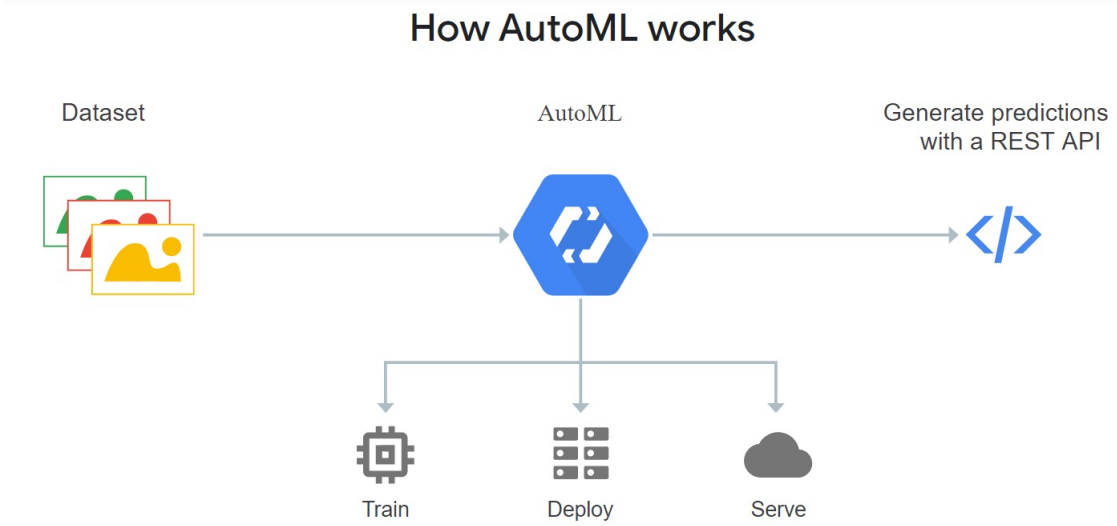
TPOT , GCP Cloud AutoML, Auto Keras, Auto SkLearn

Give me the link?

https://cloud.google.com/automl
https://automl.github.io/auto-sklearn/master/
https://epistasislab.github.io/tpot/

Auto ML





The MLOps Project

OMG! Stop, I have a question?

Multi Model Analysis

What is it?

Building Multiple models in one go and doing post-model analysis

Why do it?

Increase the performance of complex ML tasks by comparing multiple models

How to do it?

PyCaret, low-code ML library that allows you to do multi model analysis and go from preparing your data to deploying your model

Give me the link?

https://pycaret.org/guide/

Model Results

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa
0	Logistic Regression	0.7524	0.7889	0.5292	0.6919	0.597	0.4233
1	Ridge Classifier	0.7523	0	0.5178	0.7026	0.591	0.4198
2	Linear Discriminant Analysis	0.7522	0.7811	0.5231	0.6967	0.593	0.4209
3	Extreme Gradient Boosting	0.7391	0.7906	0.533	0.6568	0.585	0.3982
4	Gradient Boosting Classifier	0.7354	0.7904	0.5494	0.6383	0.5871	0.395
5	Ada Boost Classifier	0.7222	0.7576	0.5596	0.6139	0.582	0.3754
6	Random Forest Classifier	0.7204	0.7686	0.4848	0.6323	0.5446	0.3485
7	CatBoost Classifier	0.7167	0.7866	0.517	0.6095	0.5571	0.3515
8	Extra Trees Classifier	0.7094	0.7496	0.469	0.6041	0.5245	0.3215
9	Light Gradient Boosting Machine	0.7094	0.757	0.5339	0.5949	0.5595	0.3443
10	K Neighbors Classifier	0.702	0.7215	0.5073	0.6009	0.5456	0.3259
11	Naive Bayes	0.6816	0.7272	0.2719	0.5903	0.3625	0.1939
12	Decision Tree Classifier	0.6519	0.6148	0.4924	0.5024	0.4948	0.2302
13	Quadratic Discriminant Analysis	0.5772	0.5825	0.4611	0.4199	0.3621	0.092
14	SVM - Linear Kernel	0.5192	0	0.5585	0.3294	0.3521	0.0498

Model and Data Bias

What is it?

Every Model & Data Might Lie & It's important to address that.

Why do it?

Unwanted bias in datasets and machine learning models may cause irreversible damages to enterprise

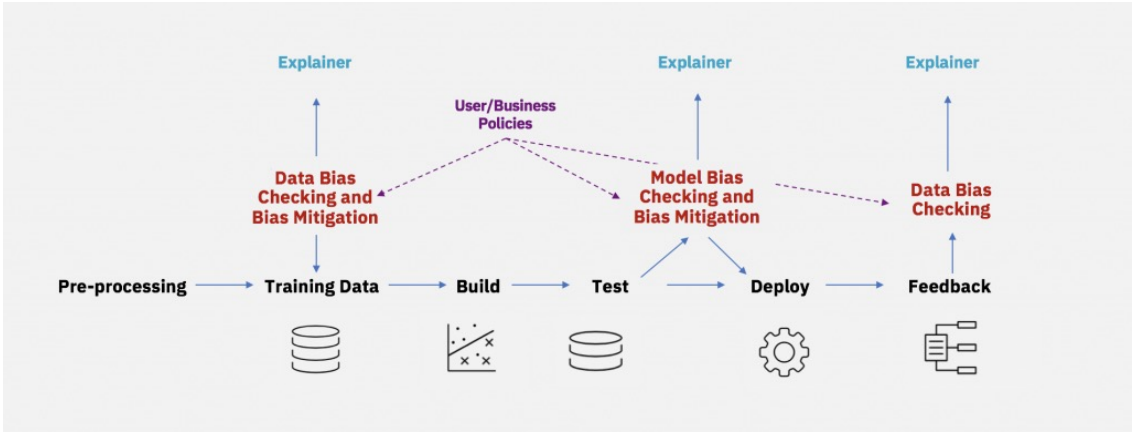
How to do it?

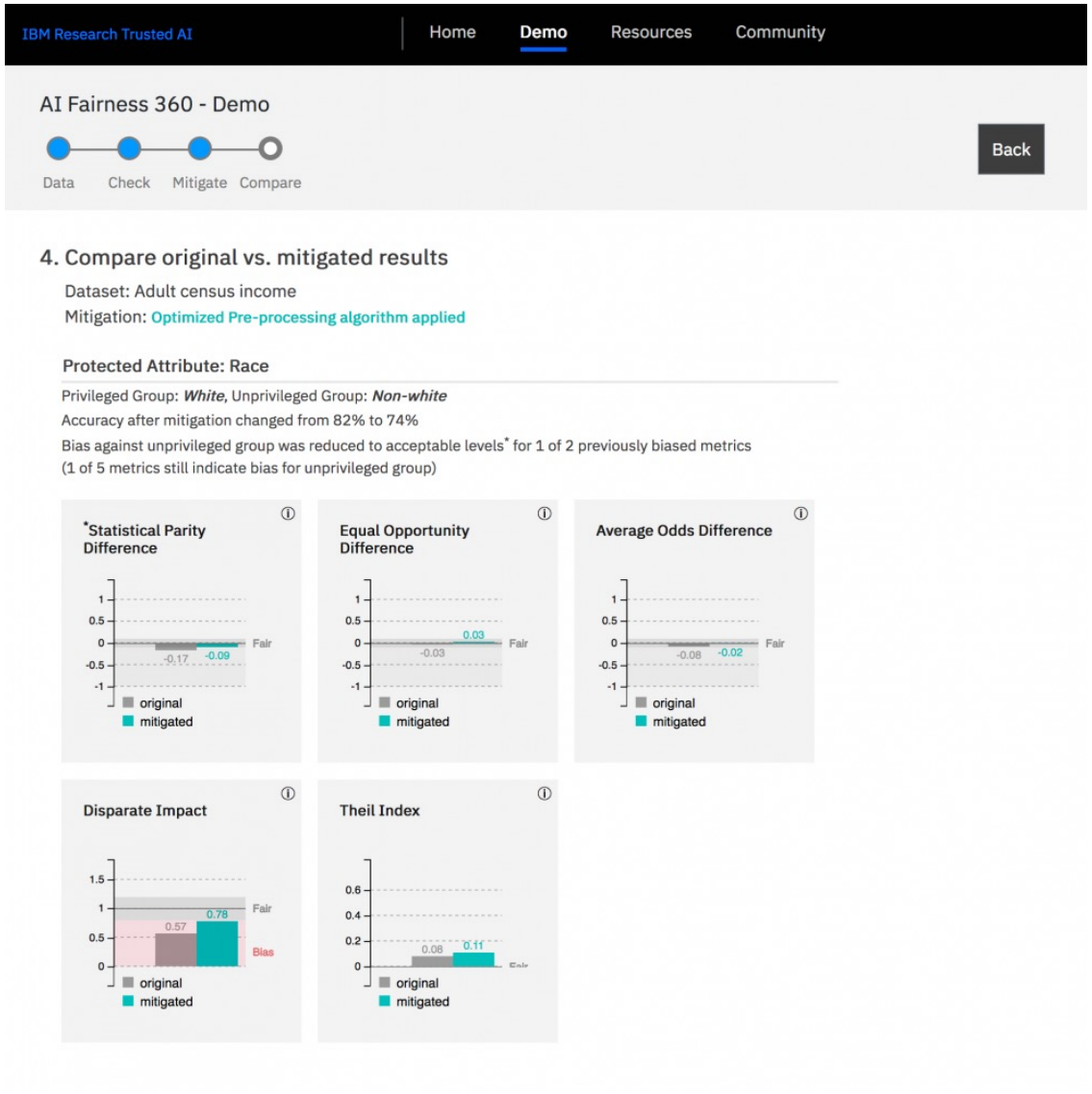
You can use IBM 360 Fairness and FairML for your data sets and Machine learning biasness

Give me the link?

https://aif360.mybluemix.net/
https://github.com/adebayoj/fairml

The MLOps Project





Distributed Data Science

What is it?

Scalable computing , Resilient process , Simple to implement and distribute data and computation over multiple CPU/GPUs

How to do it?

Use Rapid / DASK for ETL integration , Pre-Processing and ML Models

Why do it?

Enterprise play at very high scale w.r.t to data and hence ML tools need to adapt scale.

Give me the link?

https://rapids.ai/index.html

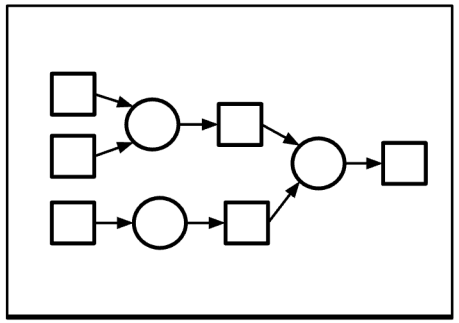
https://dask.org/

Collections
(create task graphs)

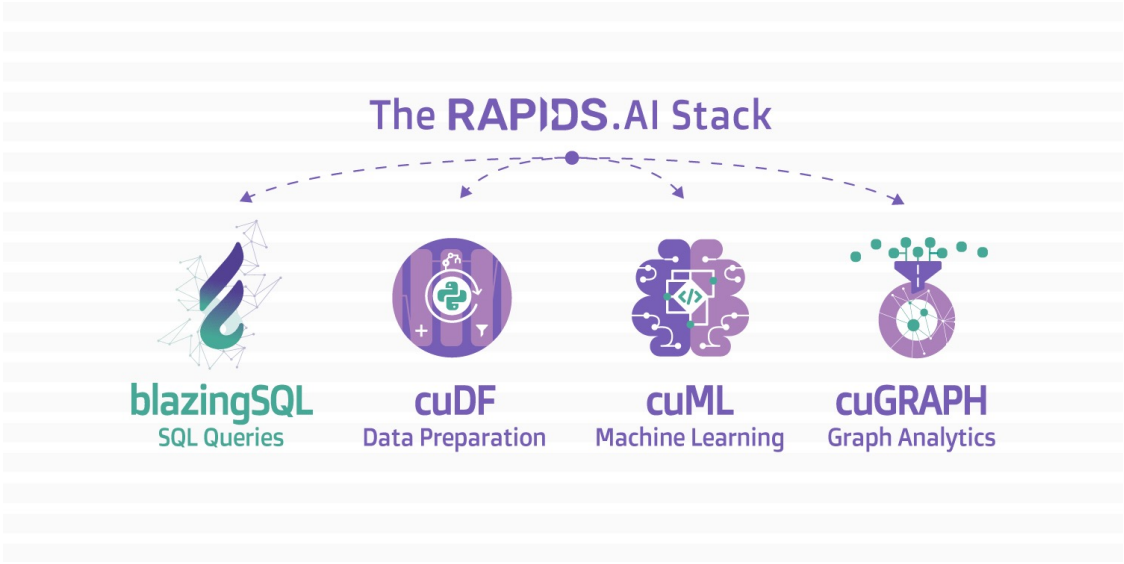
→ **Task Graph** →

Schedulers
(execute task graphs)

- Dask Array
- Dask DataFrame
- Dask Bag
- Dask Delayed
- Futures



- **Single-machine**
  (threads, processes, synchronous)
- **Distributed**



Data Interactive Apps

What is it?

Build your own interactive data apps that can run models and analysis by click.

Why do it?

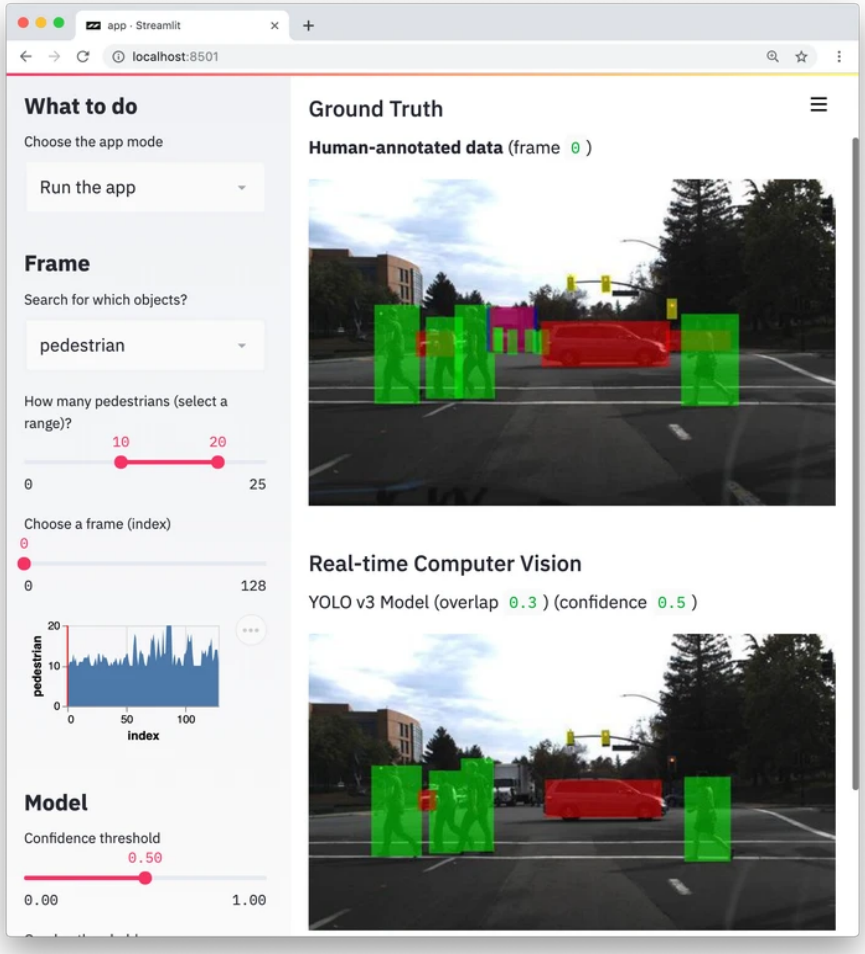
-Enterprises are looking forward to have fastest way to build custom ML tools. For easy visualization and business operation.

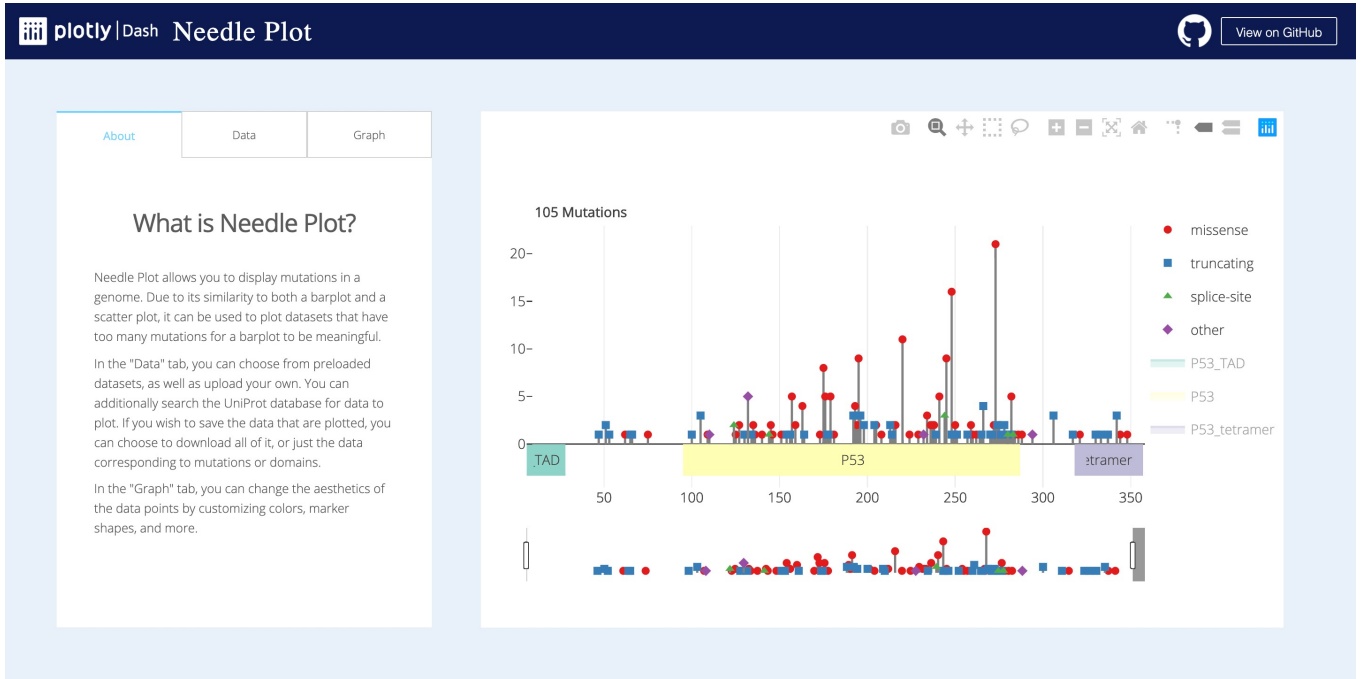
How to do it?

Streamlit/ Dash lets you create apps for your machine learning projects with deceptively simple Python scripts

Give me the link?

https://www.streamlit.io/
https://plotly.com/dash/





The MLOps Project

Thank you!

The MLOps Project