Mu Sigma
Chicago, IL
Bangalore, India
www.mu-sigma.com

Thursday Learning Hour

28th July 2022

Beyond .Fit()

Navigating the robust solution

Democratizing ML series – Session 16

Prabakaran Chandran
Decision Scientist & Apprentice Leader

What are we going to Learn today ?

- Recap - .fit()
- Models that are not just limited with .fit()
- Gaps in Most of the ML Solutions in the delivery
- Keep MLOPs as the SOP
- Build visibility through Model Explainability
- Quantify the Model's Uncertainty
- Monitor your Model's health
- Package your Models
- ML Ethics
- Don't Stop Learning – Keep upgrading your Models

.Fit() brings the Confidence on your Solution in the early stage – Let's recap the typical ML Development Activities

Mu Sigma

- muPDNA (Problem Definition)
- Data Discovery
- EDA and Hypotheses testing

```
from sklearn.linear_model import LinearRegression
model = LinearRegression(fit_intercept=True)

model.fit(x[:, np.newaxis], y)

xfit = np.linspace(0, 10, 1000)
yfit = model.predict(xfit[:, np.newaxis])
```

- Deploy as Notebook Pipelines
- Deploy as Web Apps (Real time)
- Post Deployment Analysis

Let's Not underestimate…..! Most of the time , **Feature Engineering and Hyperparameter tuning** are the real trump cards

Mu Sigma

Discovery

- muPDNA (Problem Definition)
- Data Discovery
- EDA and Hypotheses testing

Deployment

- Deploy as Notebook Pipelines
- Deploy as Web Apps (Real time)
- Post Deployment Analysis

Feature Eng.

```
from sklearn.linear_model import LinearRegression
model = LinearRegression(fit_intercept=True)

model.fit(x[:, np.newaxis], y)

xfit = np.linspace(0, 10, 1000)
yfit = model.predict(xfit[:, np.newaxis])
```

Hyperparameter tuning

Addressing the Gap is all we need

Mu Sigma

Where do we spent our more time ?

- Data Preparation
- EDA (Not Really – But Ideally we should)
- Model Development and Improvement (In a Closed Space : Not Much Exploration)

What do we focus much ?

- Getting better performance (Mostly by considering wrong metrics)

What do we consider as success ?

- Accuracy : (is all we wanted
- What about Optimization / Computation Efficiency then?

- **Identified new ML Solutions**
- **Explainable Models**
- **Models under Our Radar**
- **Reliable Predictions**
- **Continuous Training**
- **Easier Portability**
- **Enhanced Latency**
- **Ensured Ethics**

Models that needs more handcrafting and solve unique problems – ML is not at all a Single Layered Solution

Mu Sigma

Measuring the real Cause	- Models that can capture the real causal impact between the variables
Handling Unstructured Data	- Models that can handle multi modalities
Probabilistic – Yet Realistic	- Models that can be trained in Bayesian way
Hyper Personalization	- Models that can bring personalization and handle cold start problem

Trade off between **Exploitation and Exploration** of Machine Learning Solutions - Unique Solutions to Show Off ☺

Mu Sigma

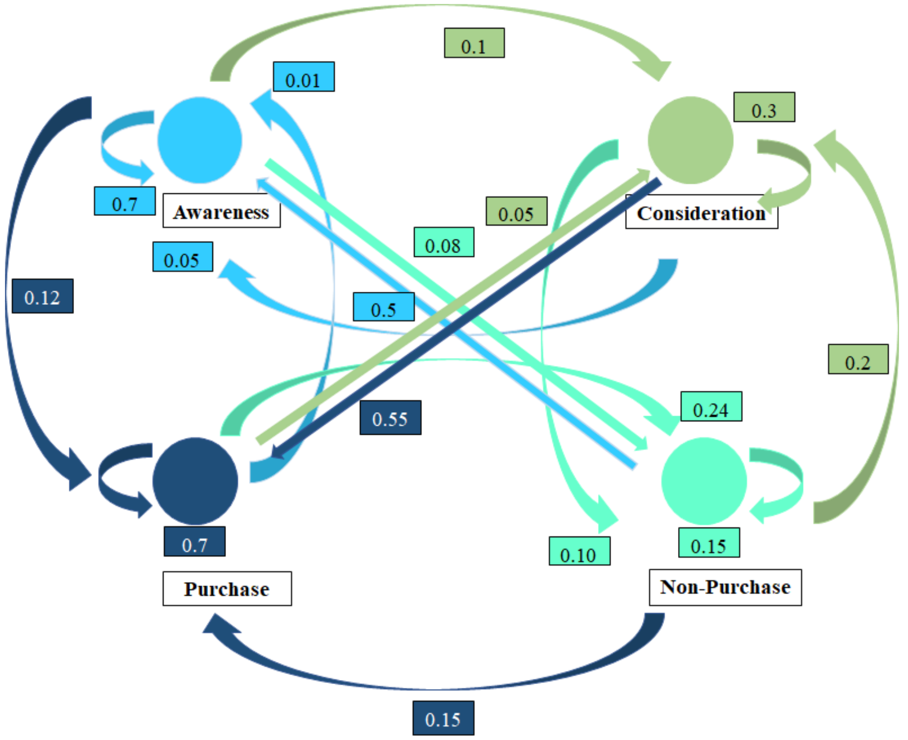
Fulfilment ←→ **Growth**

Exploitation	**Equilibrium**	**Exploration**
XGBoost		Graph Machine Learning
CARTS and Simple straight forward Linear Regression		Causal ML and Probabilistic Models
Very Old ways to solve Unstructured data problems		Transformers and Self supervision Models
Slower adaption of Cutting edge technologies		Graphs / Digital twins / Reinforcement Learning

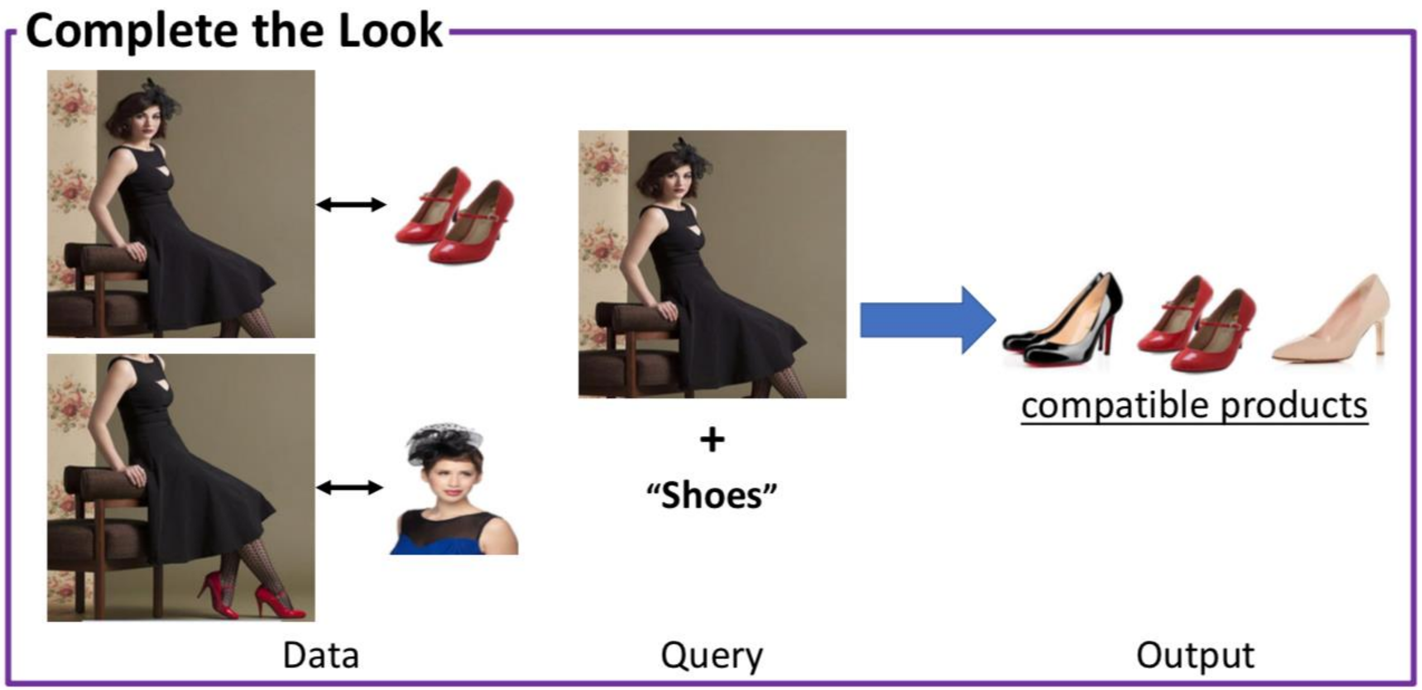
Trade off between **Exploitation and Exploration** of Machine Learning Solutions - Unique Solutions to Show Off ☺

Mu Sigma

Markov Chains for Marketing analytics



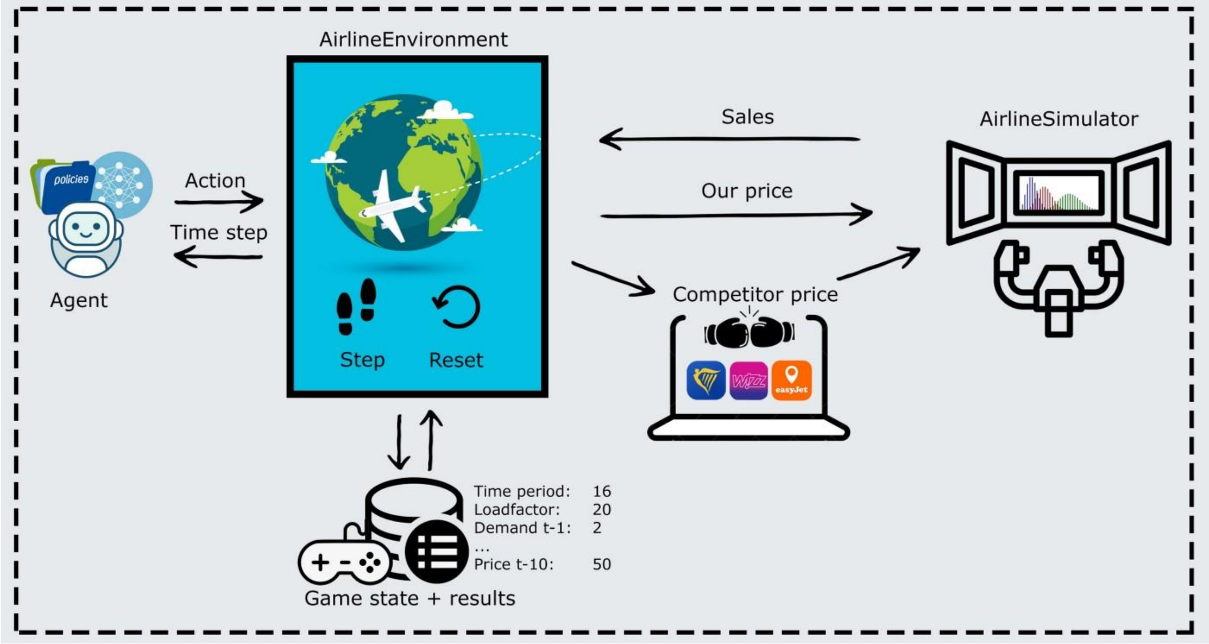
Vision and Language Modelling for Hyper personalization



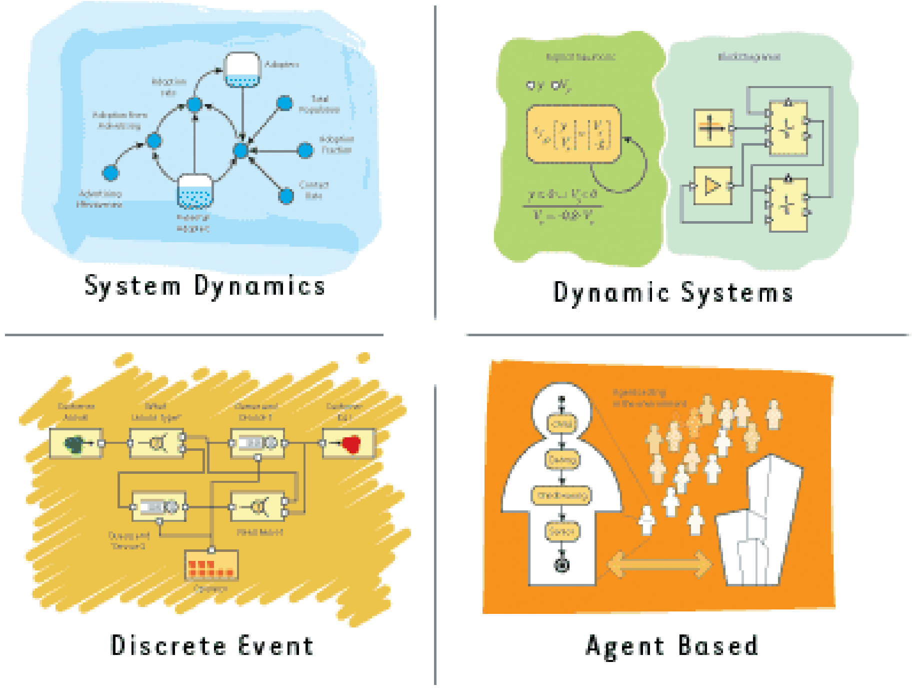
Trade off between **Exploitation and Exploration** of Machine Learning Solutions - Unique Solutions to Show Off ☺

Mu Sigma

Reinforcement Learning for Dynamic Pricing



Twinning through Simulations



Focusing the **Post Model Development Activities** to enhance the ML Solutions

Mu Sigma

Packaging Solutions

ML Ethics

Model Health and Drift Analysis

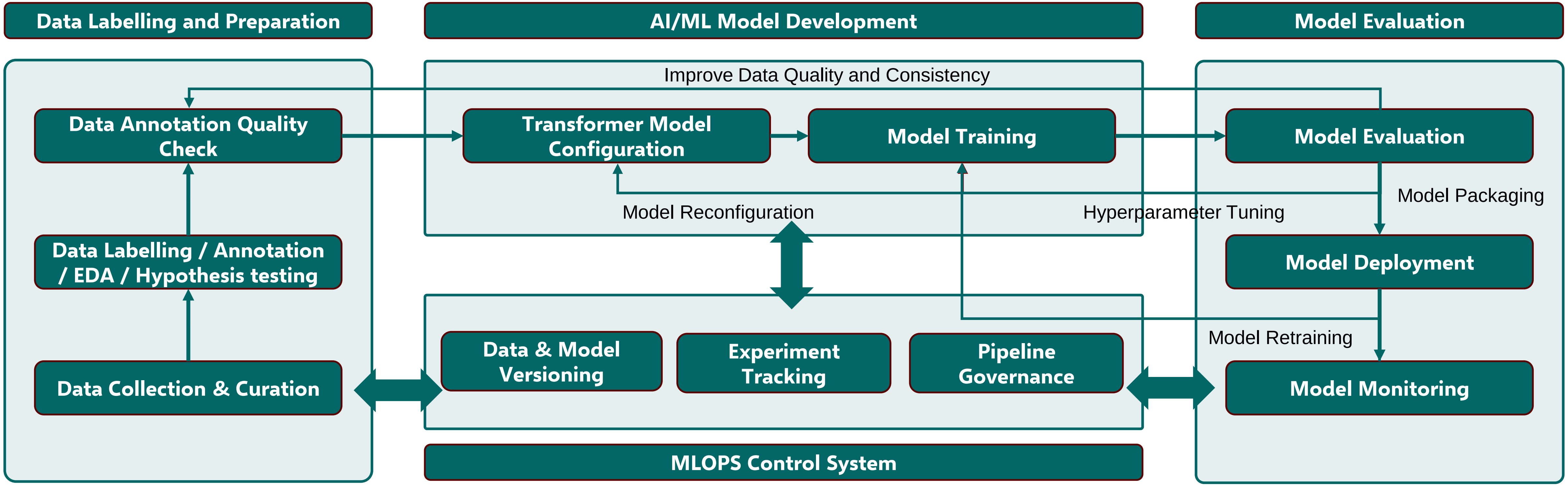
Model Calibration and Uncertainty Quantification

Model Explainability and Interpretability

Machine Learning Operations Practice

Model Development

MLOps as the Standard SOP to build ML Solutions – Enabling Clients to adapt/explore MLOps powered Solutions



Tools and Frameworks Required

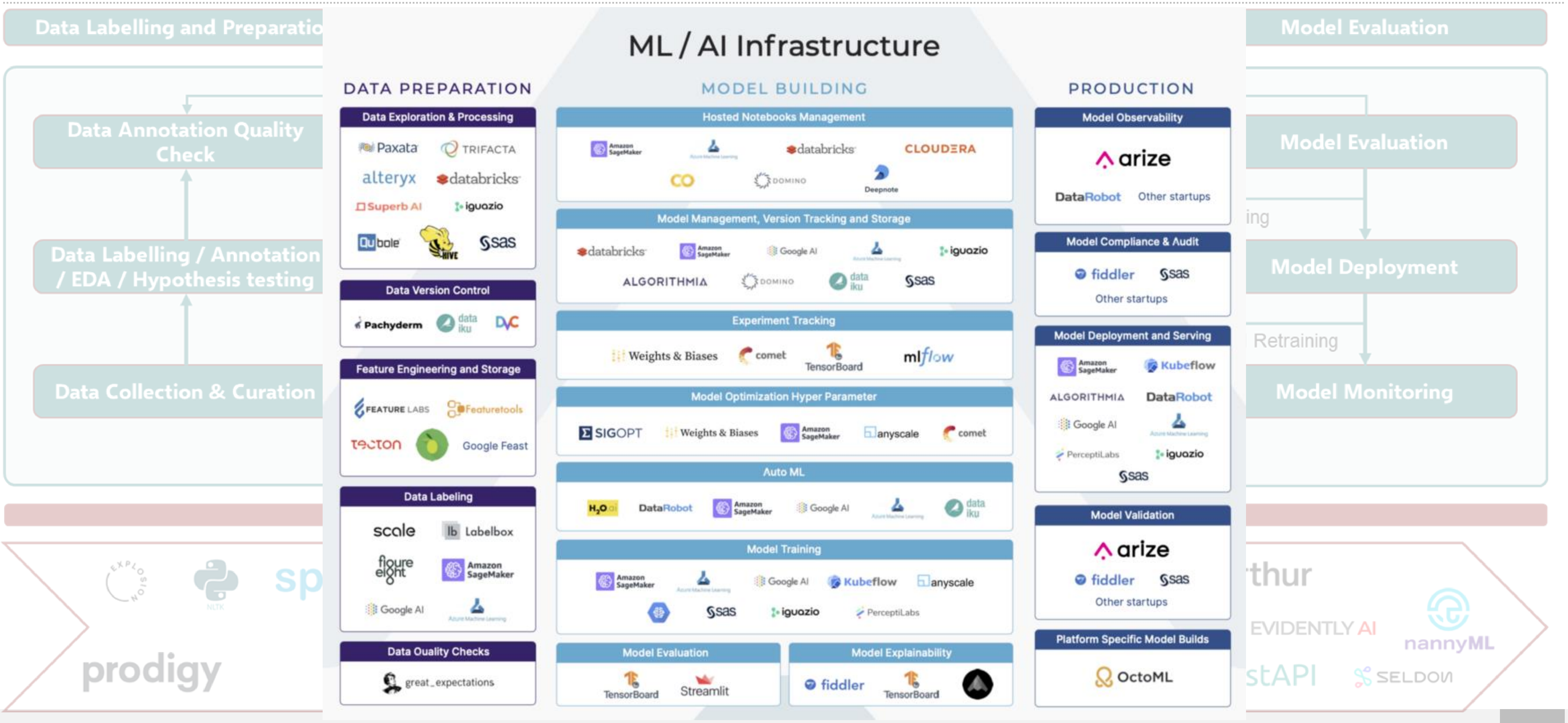
Explosion, NLTK, spaCy, prodigy

H2O.ai, Transformers, DVC, PyTorch, scikit learn, comet, Weights & Biases

Arthur, Evidently AI, nannyML, FastAPI, Seldon

Explore the MLOPs Landscape

Mu Sigma



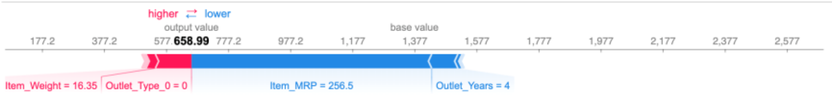
Building **Visibility through Model Explainability** – Make your Black box Models into White/Glass Box Models

Linear Models

$$\underbrace{Y_i}_{\text{Dependent Variable}} = \overbrace{\beta_0}^{\text{Constant/Intercept}} + \underbrace{\beta_1}_{\text{Slope/Coefficient}} \overbrace{X_i}^{\text{Independent Variable}}$$

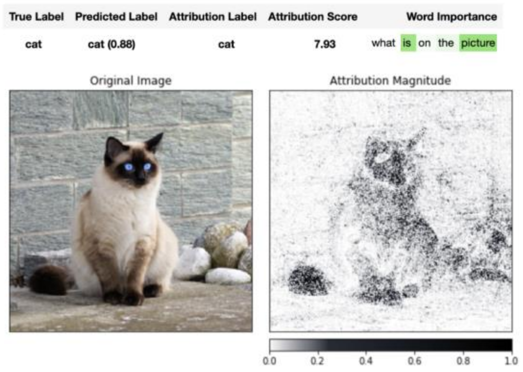
- Interpreting the Association not the Causation
- But We can use Causal ML Practices

CARTS

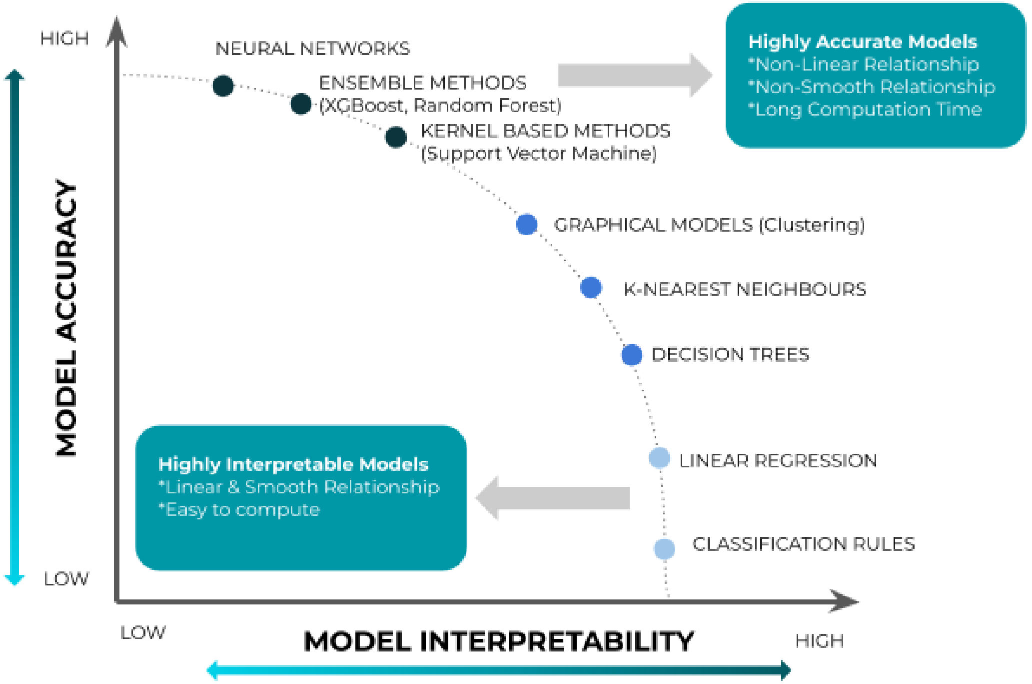


- Shapely / LIME Values

Deep Learning Models



Text Contributions: 7.93
Image Contributions: 3.12
Total Contribution: 11.05



Why Uncertainty matters ?

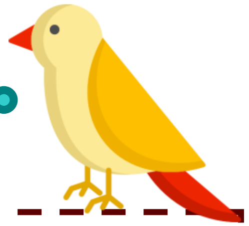
Decision Making Process should be supported with a Prediction with a level of Confidence

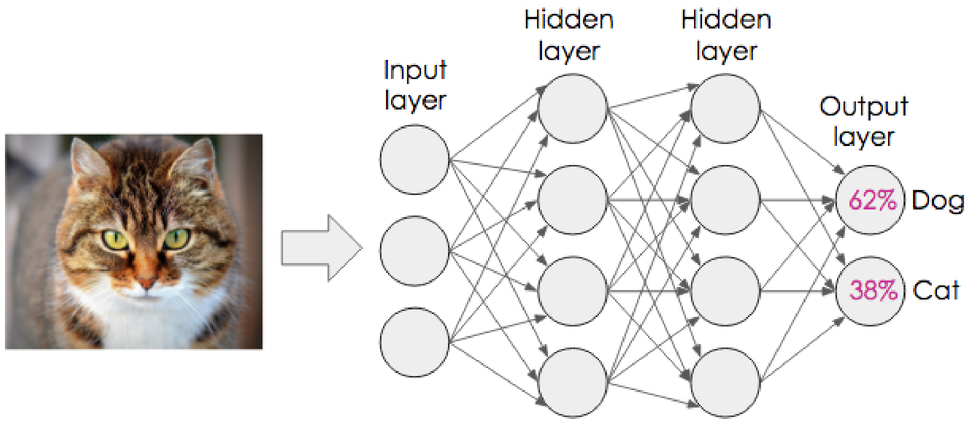
You may tell , We have Error metrics to measure accuracy .

But all of them are global level of confidence metrics.
So, it is necessary to look for

“Local Levels of Confidence at the Sample Level “

Hey There!
I am Chittu kuruvi!
(Lil Sparrow)
Will I be classified
as Cat/Dog





It should be
100% Cat :(

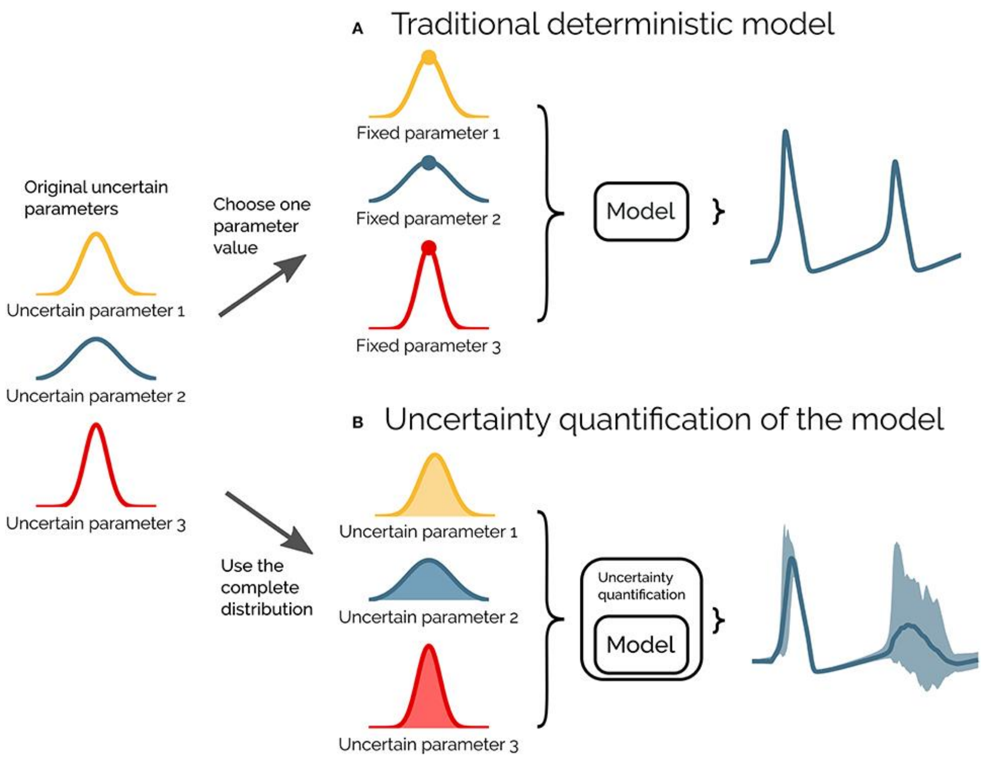
What If , the input image is a Bird
?

Will it be classified as “Dog” or
“Cat” ?

YES, the Neural network still predicts as “Dog/Cat” with a high probability score on unseen images known as *“out-of-distribution”* samples.

Model Calibration and Uncertainty Quantification to ensure confidence and reliability of your models

{X1,X2,X3….Xn} → **Any ML Model** → Yp (Point Estimate)



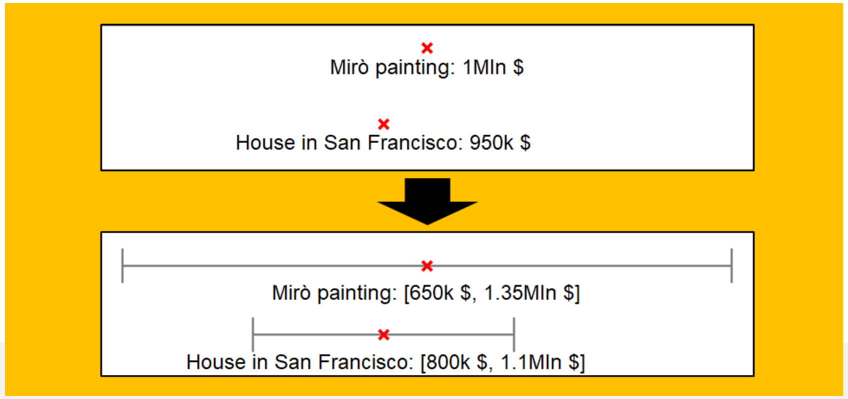
Calibration :

Calibrated probabilities means that the probability reflects the likelihood of true events.

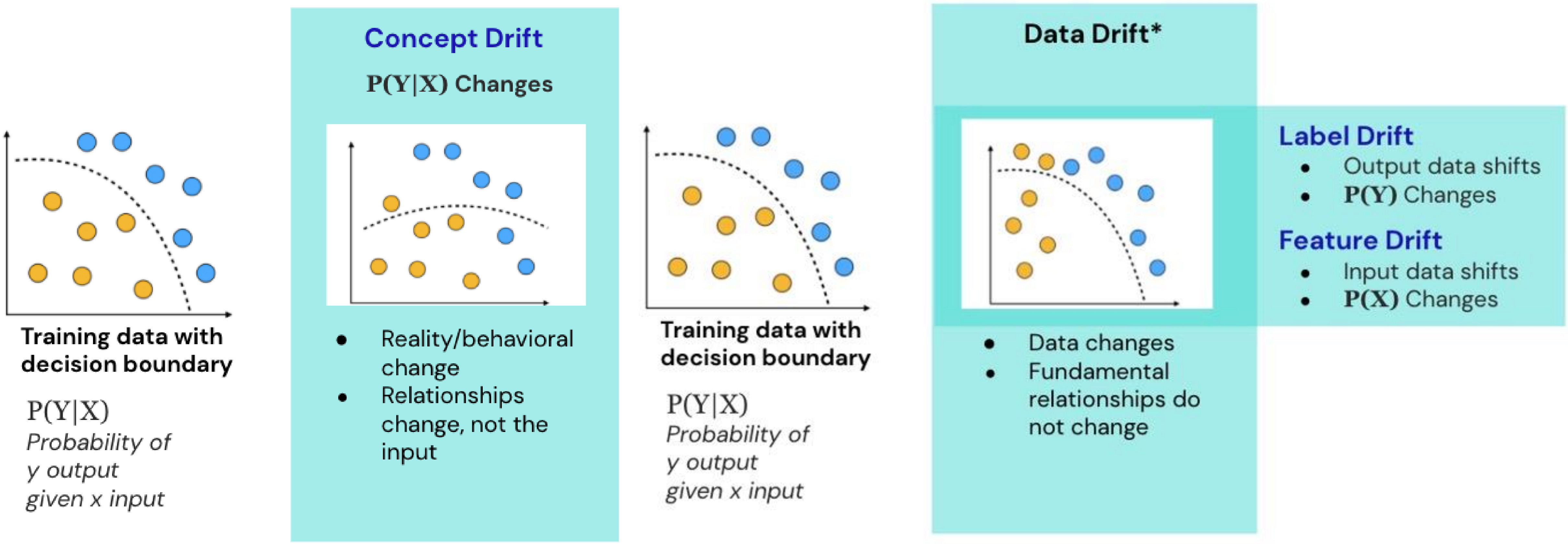
- Plat Scaling
- Isotonic Regression

Uncertainty Quantification :

Intervals around Prediction – Not the typical confidence interval

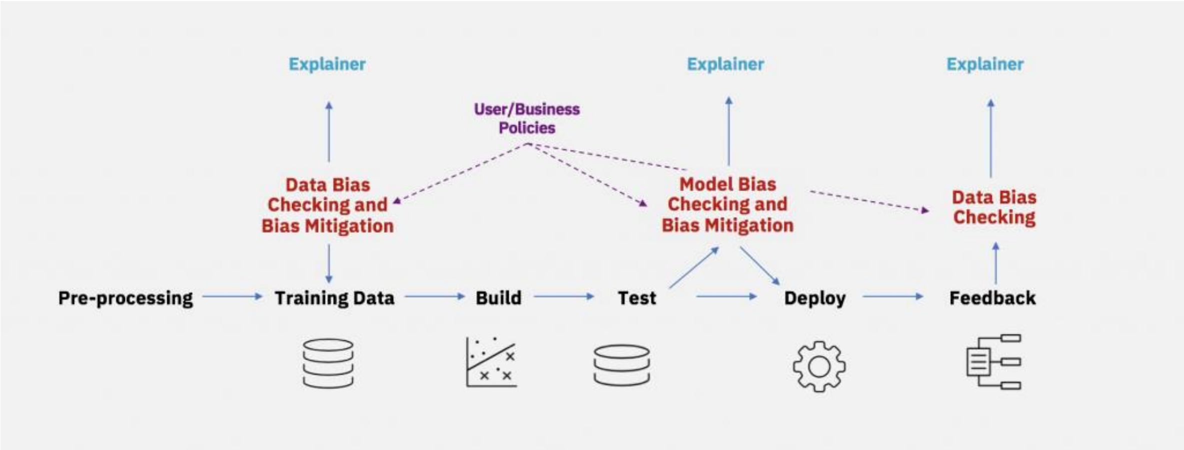


Monitoring to ensure robust ML Pipeline – Measure Post Development Performance and Drift Metrics



- Many Statistical and Information theory related metrics to measure the drifts
- Entropy , Wasserstein Distance , Kolmogorov Smirnov test , Shannon Entropy , Chi Square test

Ensure ML Ethics - Leap into **Fairness, Bias and Model Privacy**



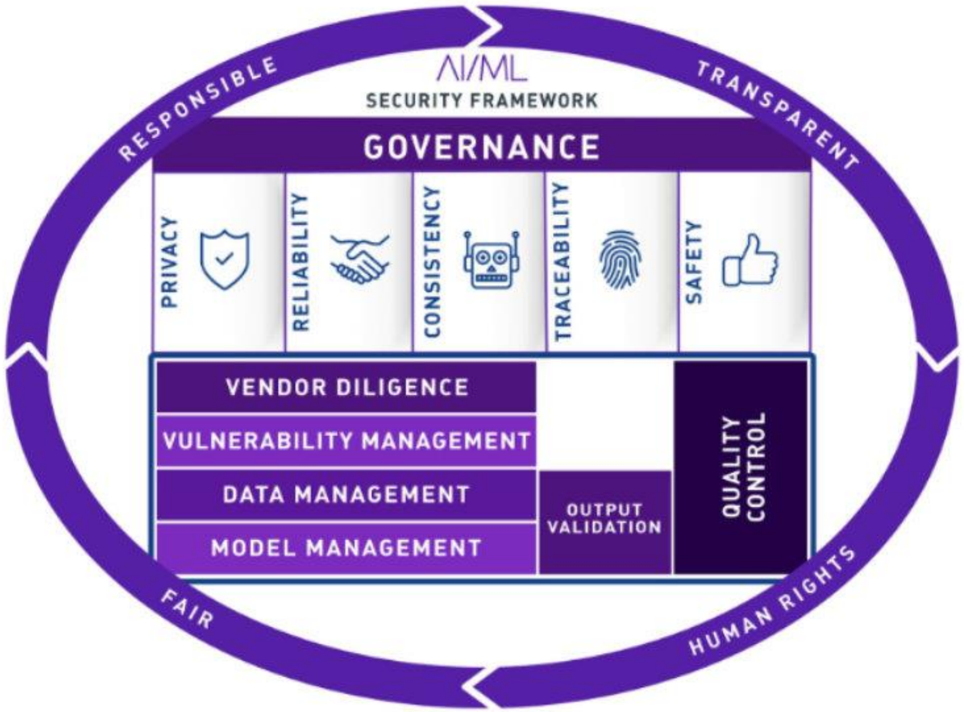
Critical solutions in Fraud / Finance / Healthcare domain needs the well balanced Fairness and Privacy

Responsible AI/ML Models

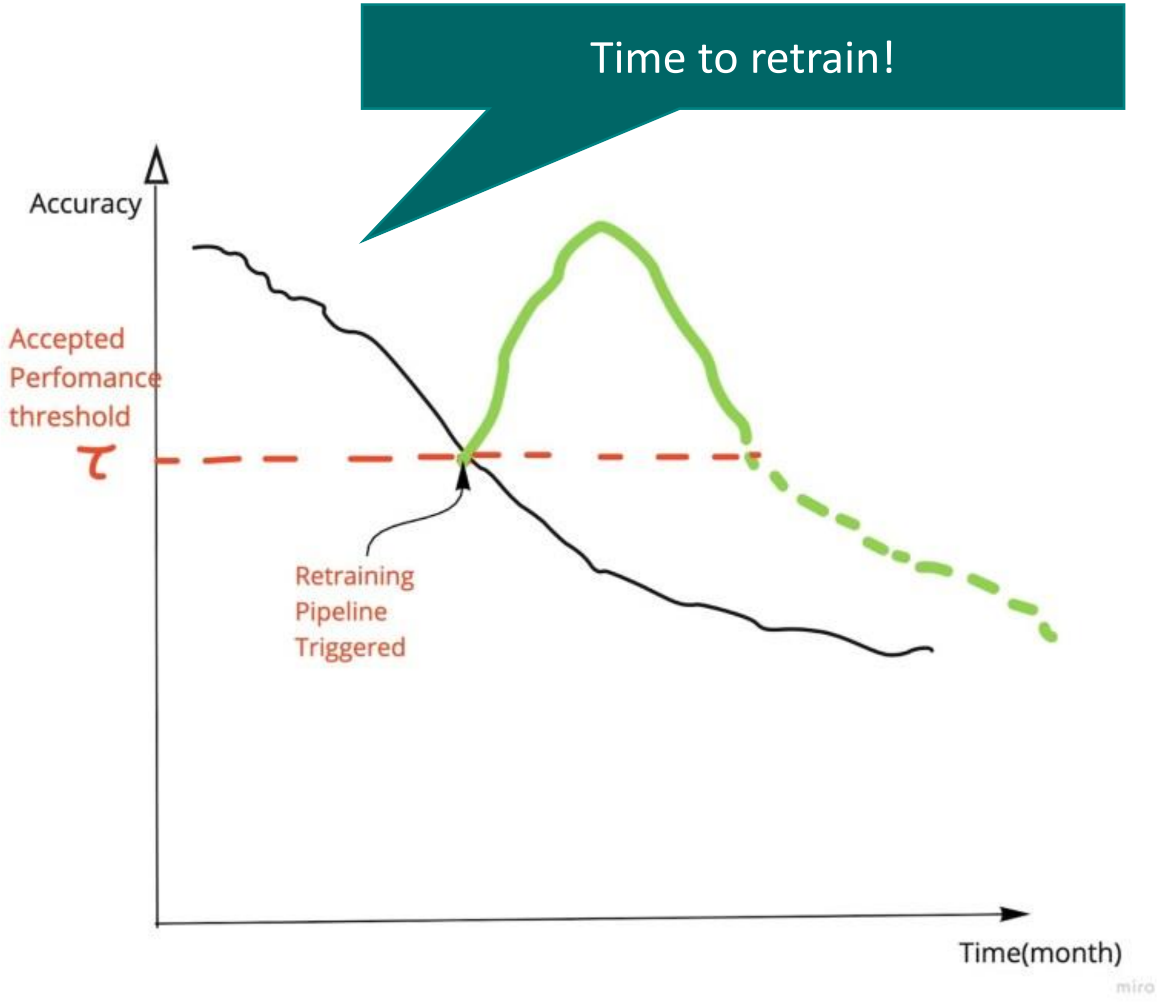
Explainable AI/ML Models

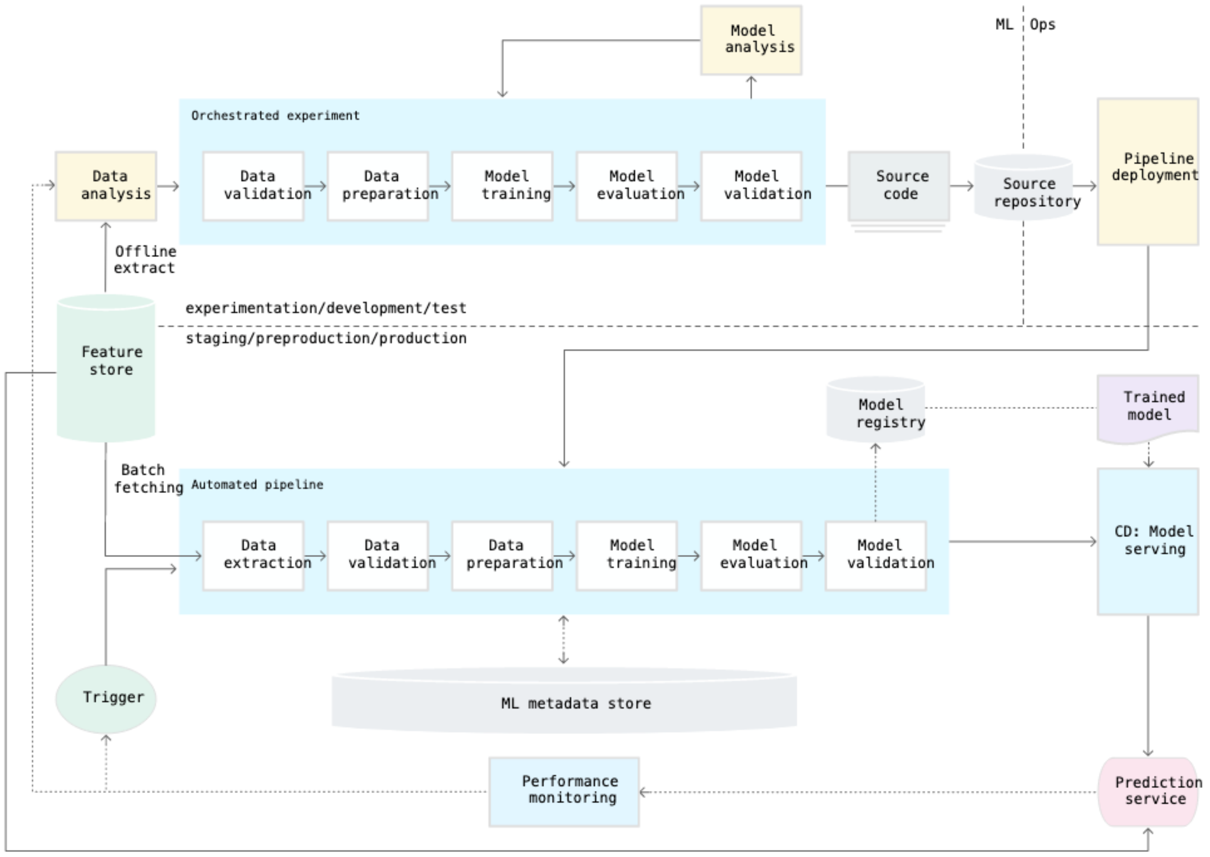
Reusable AI/ML Models

Orphaned AI / ML Models



Continuous Learning along with the Existing Continuous Integration and Continuous Deployment





Mu Sigma
Chicago, IL
Bangalore, India
www.mu-sigma.com

Questions….?

Mu Sigma
Chicago, IL
Bangalore, India
www.mu-sigma.com

Looking for ML Solutions/Help…?

Reach out to me!……. :)

Mu Sigma
Chicago, IL
Bangalore, India
www.mu-sigma.com

Thanks for joining…!

Mu Sigma
Chicago, IL
Bangalore, India
www.mu-sigma.com

Appendix…