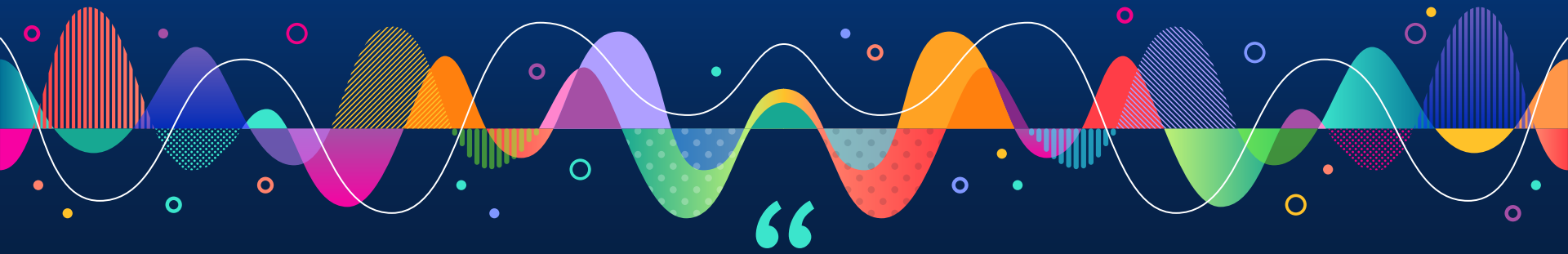


Artificial Intelligence needs Backend & DevOps to reach the real-world

Nerea Luis @sailormerqury · SNGULAR Data&AI

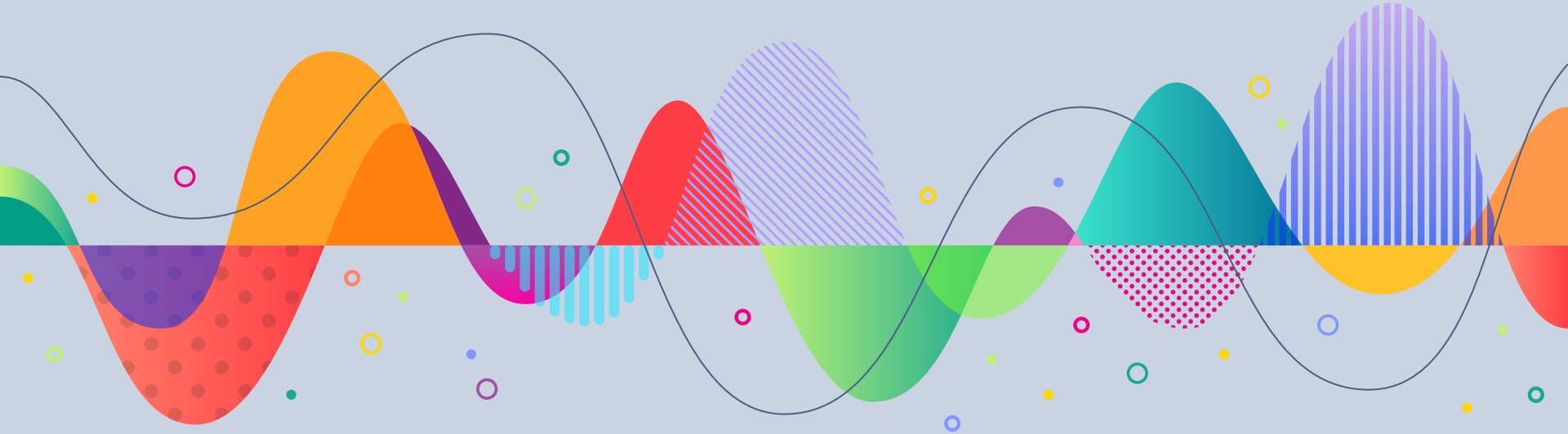
Nerea Luis (nerealuis.es)

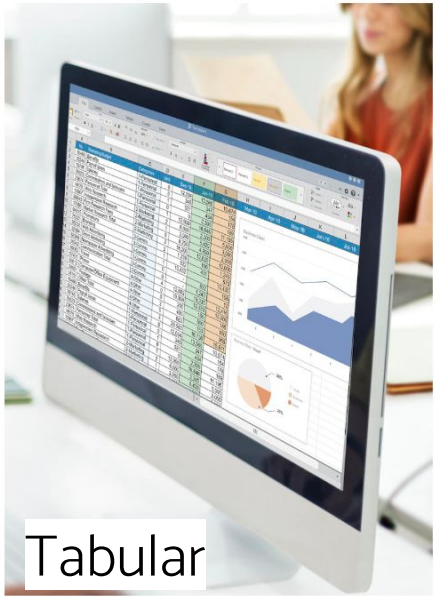
Artificial Intelligence lead @ SNGULAR

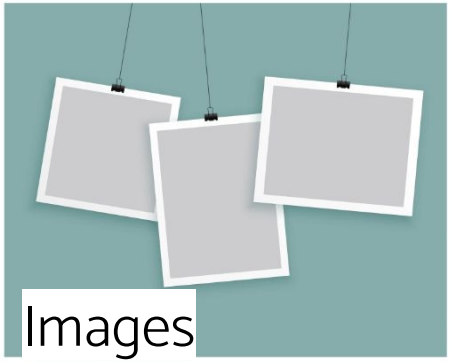
PhD Artificial Intelligence

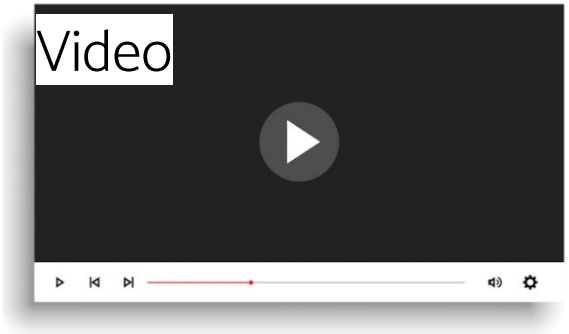
Sci-comm, women-in-tech advocate

1. The Data-Scientist work



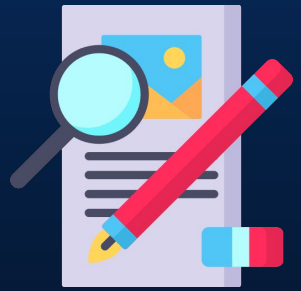






Audio

Documents



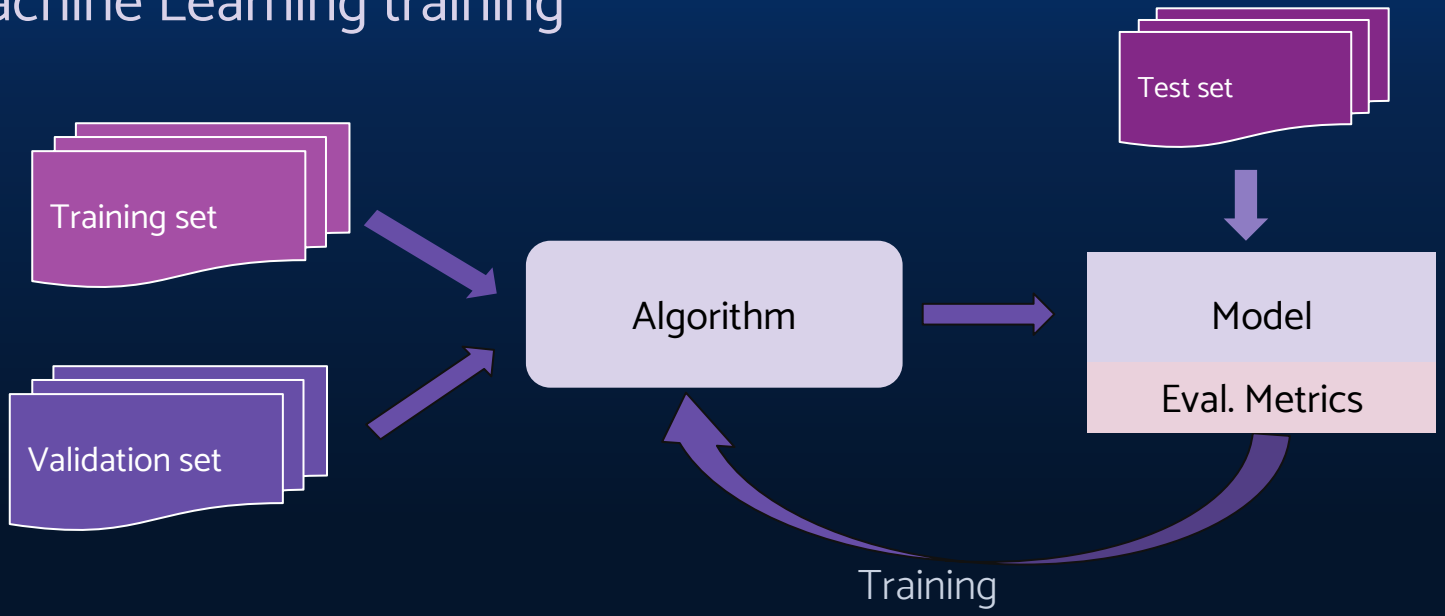
Exploring data sources

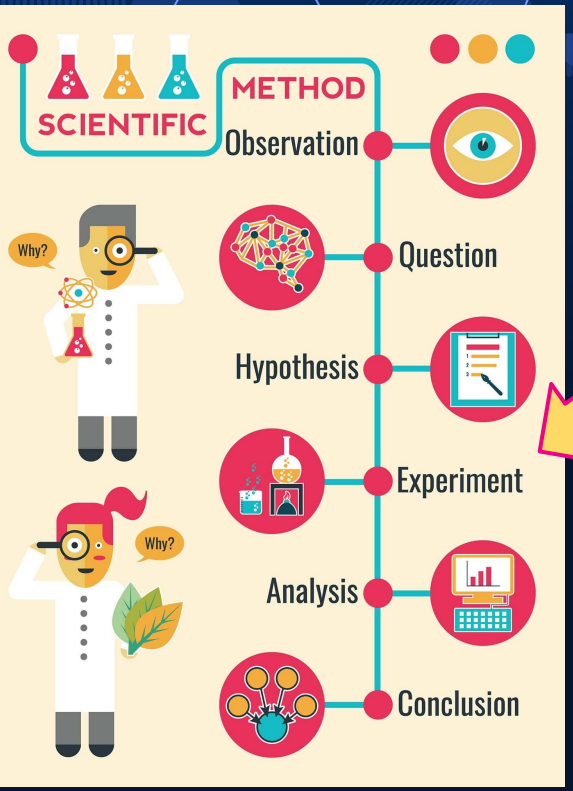
Dataset preprocessing

Exploratory Data Analysis

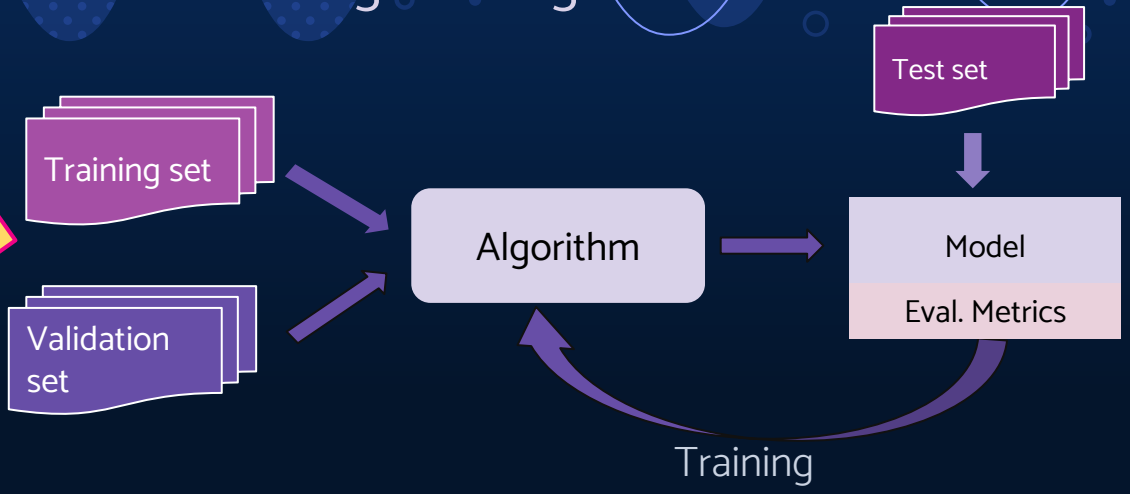
Data labeling,
data aggregation

Machine Learning training



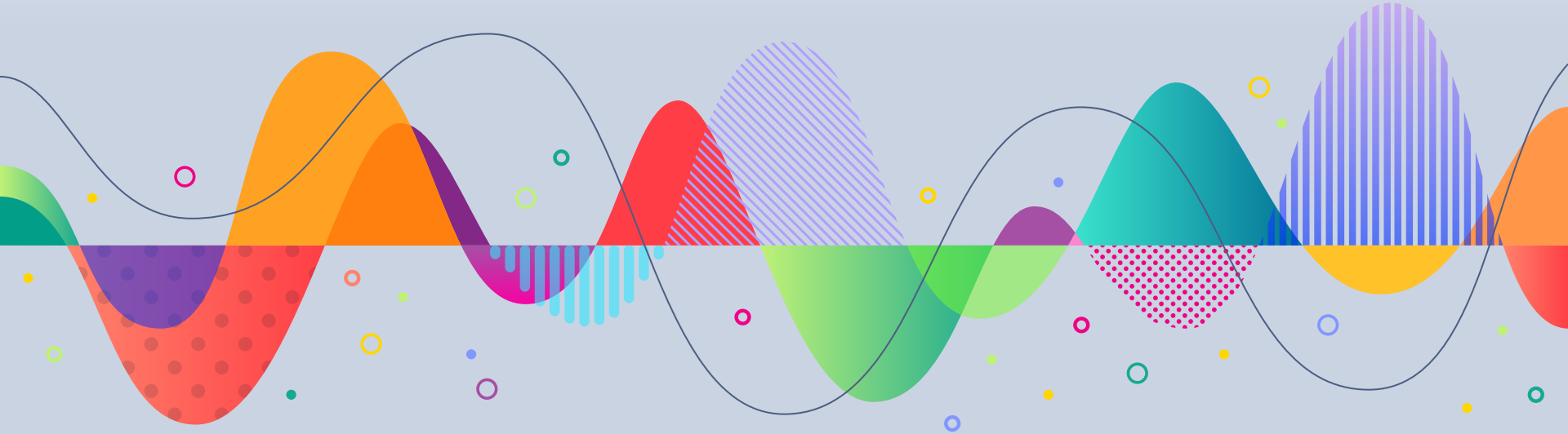


Machine Learning training



To be repeated **several times**

2. I already have an AI model. What's Next?



2015
830 cites

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.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
{ebner,vchaudhary,mwyoung,jfcrespo,dennison}@google.com
Google, Inc.

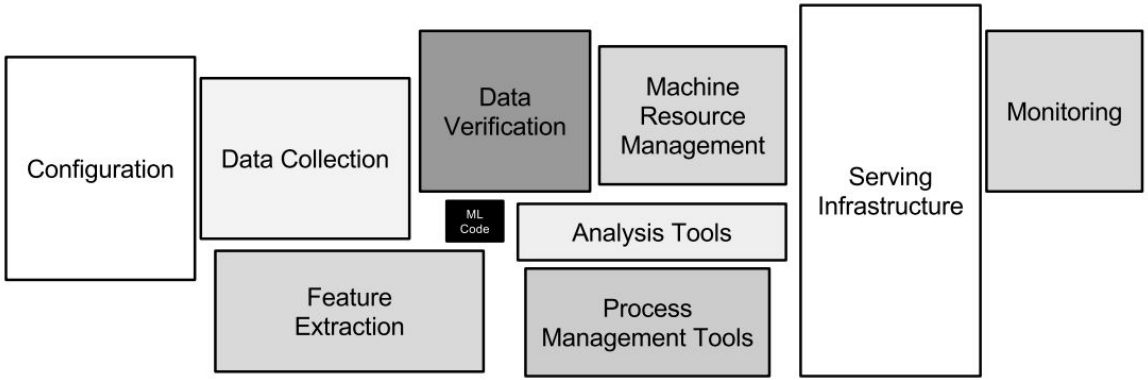
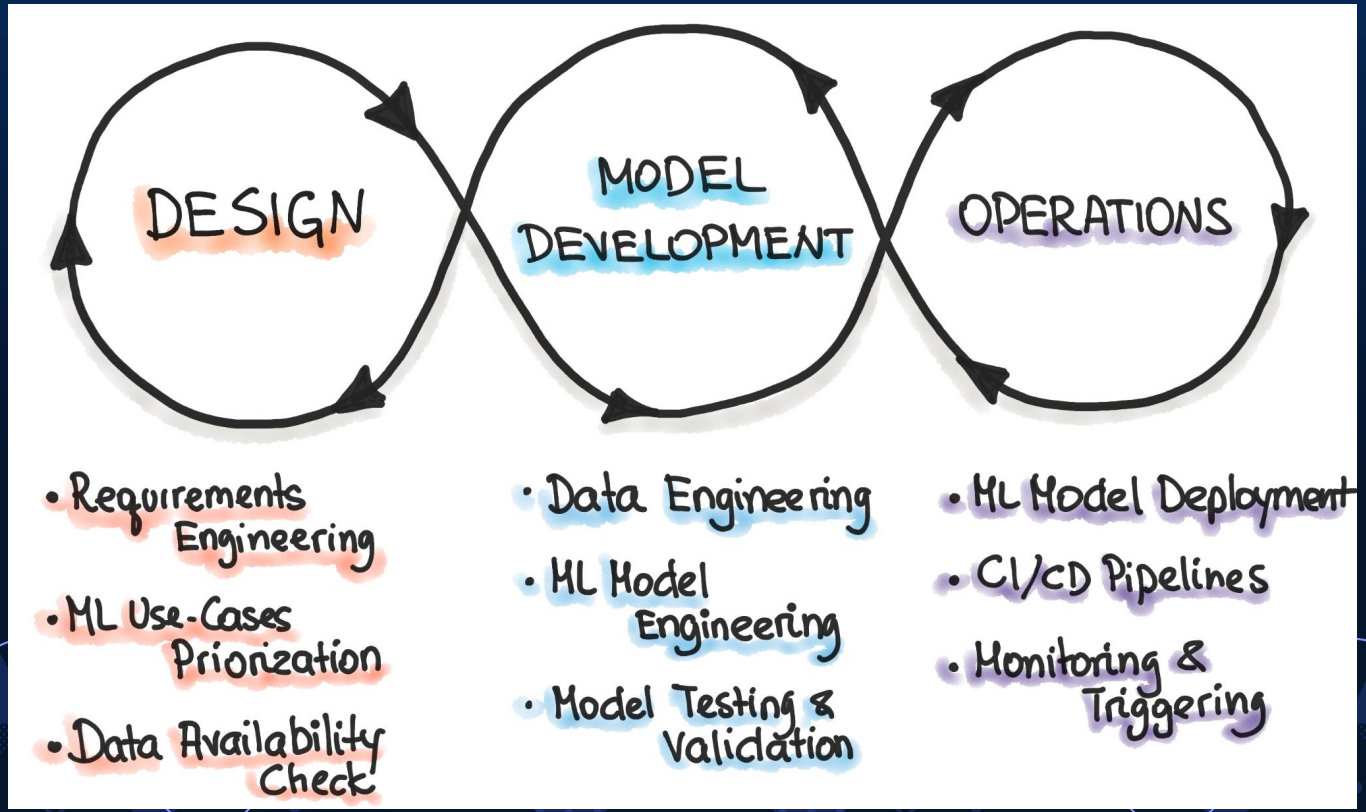


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Machine Learning “Operations”

https://ml-ops.org/



CI

Continuous Integration

CI is no longer only about testing and validating code and components, but also testing and validating data, data schemas, and models.

CD

Continuous Delivery

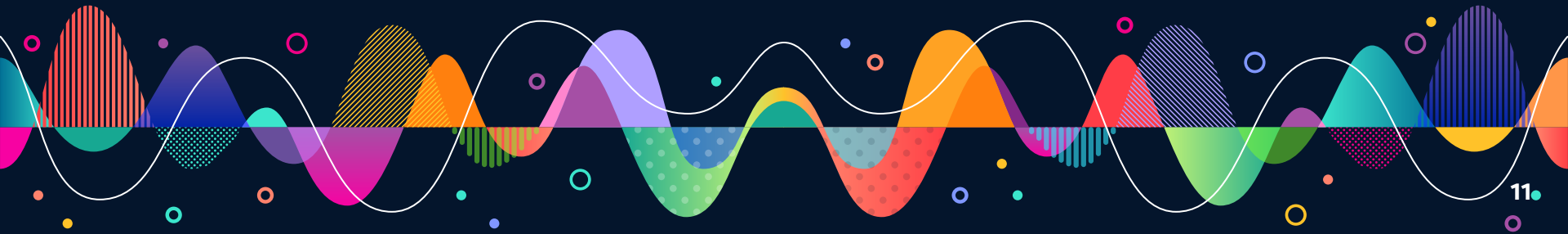
CD is no longer about a single software package or a service, but a system (an ML training pipeline) that should automatically deploy another service (model prediction service).

CT

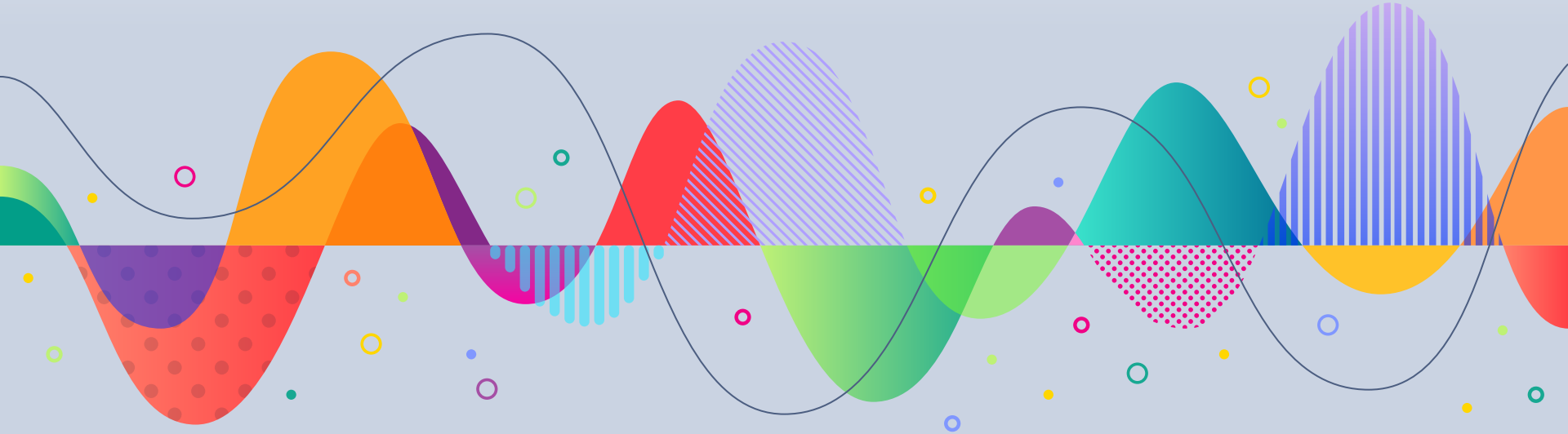
Continuous Training

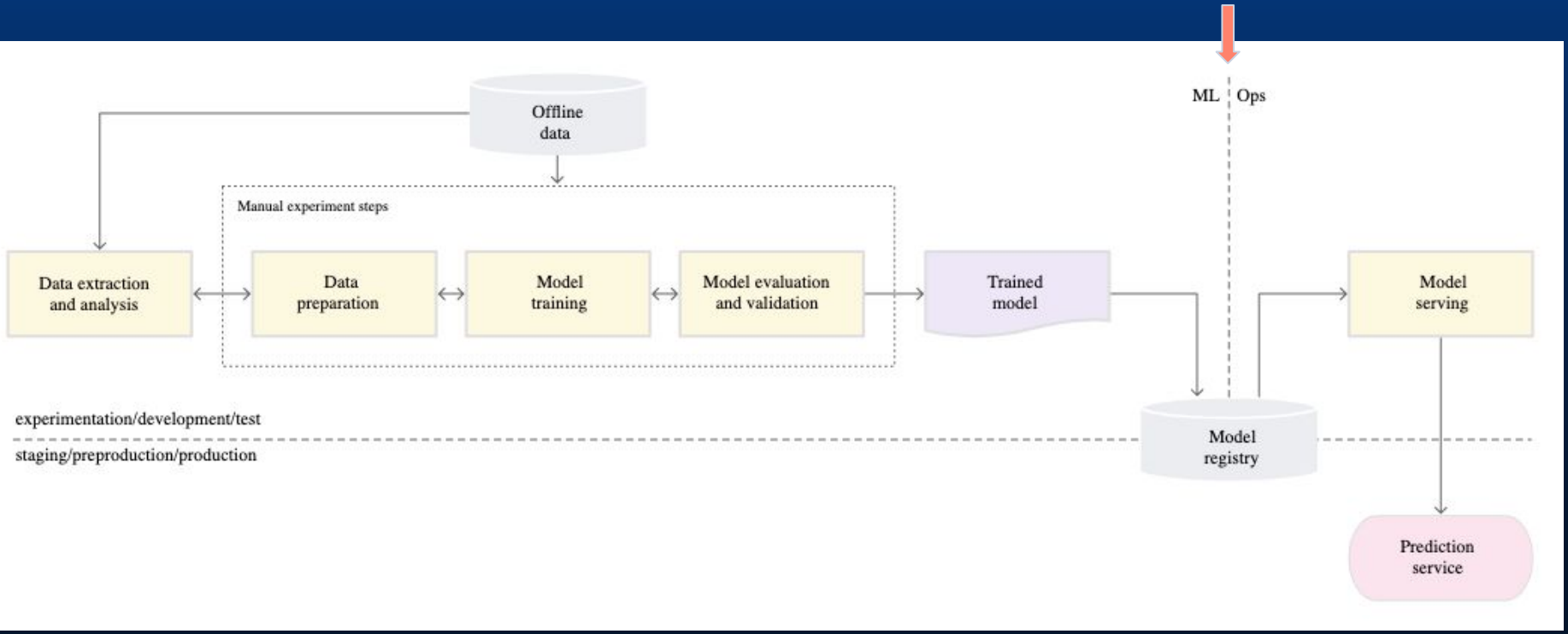
CT is a new property, unique to ML systems, that's concerned with automatically retraining and serving the models.

Reference

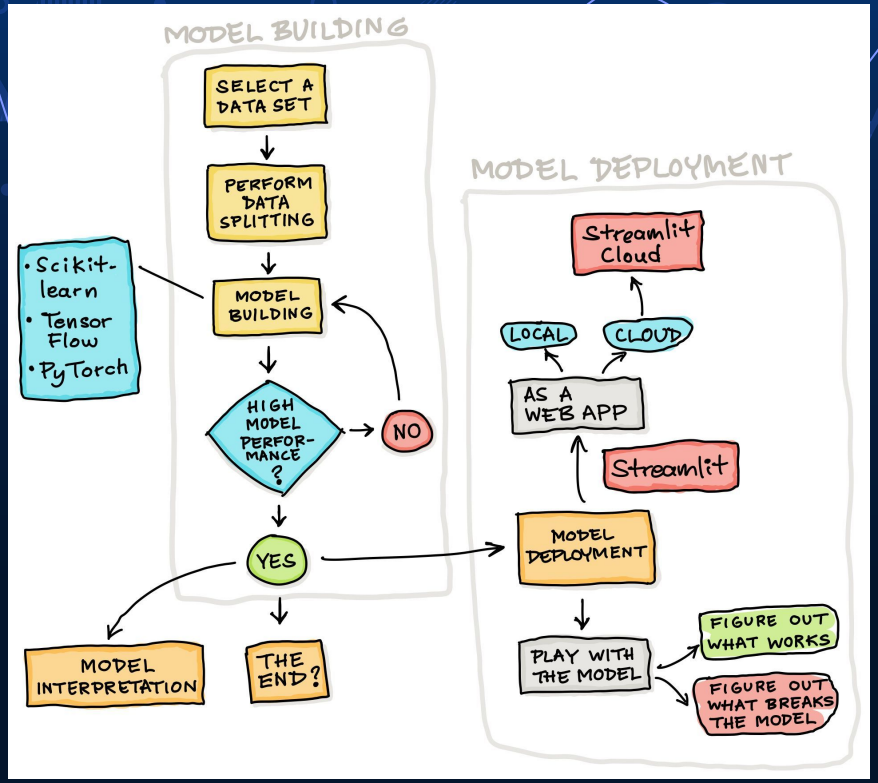


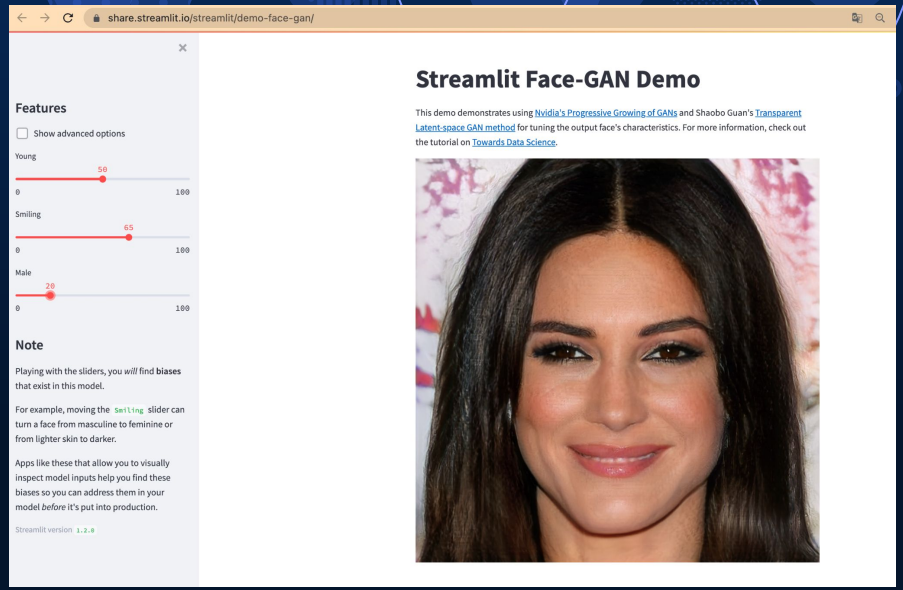
3. Accelerating AI in production




ML
Ops
Offline data
Manual experiment steps
Data extraction and analysis
Data preparation
Model training
Model evaluation and validation
Trained model
Model serving
experimentation/development/test
staging/preproduction/production
Model registry
Prediction service

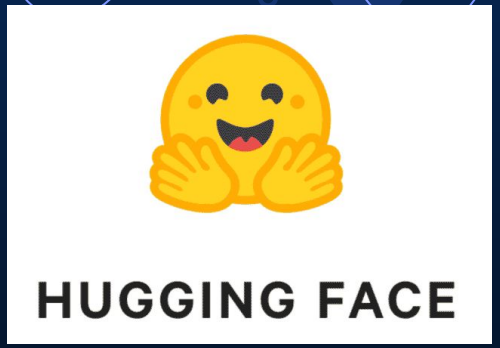
Level 1





Let's work on some GUI

Backend



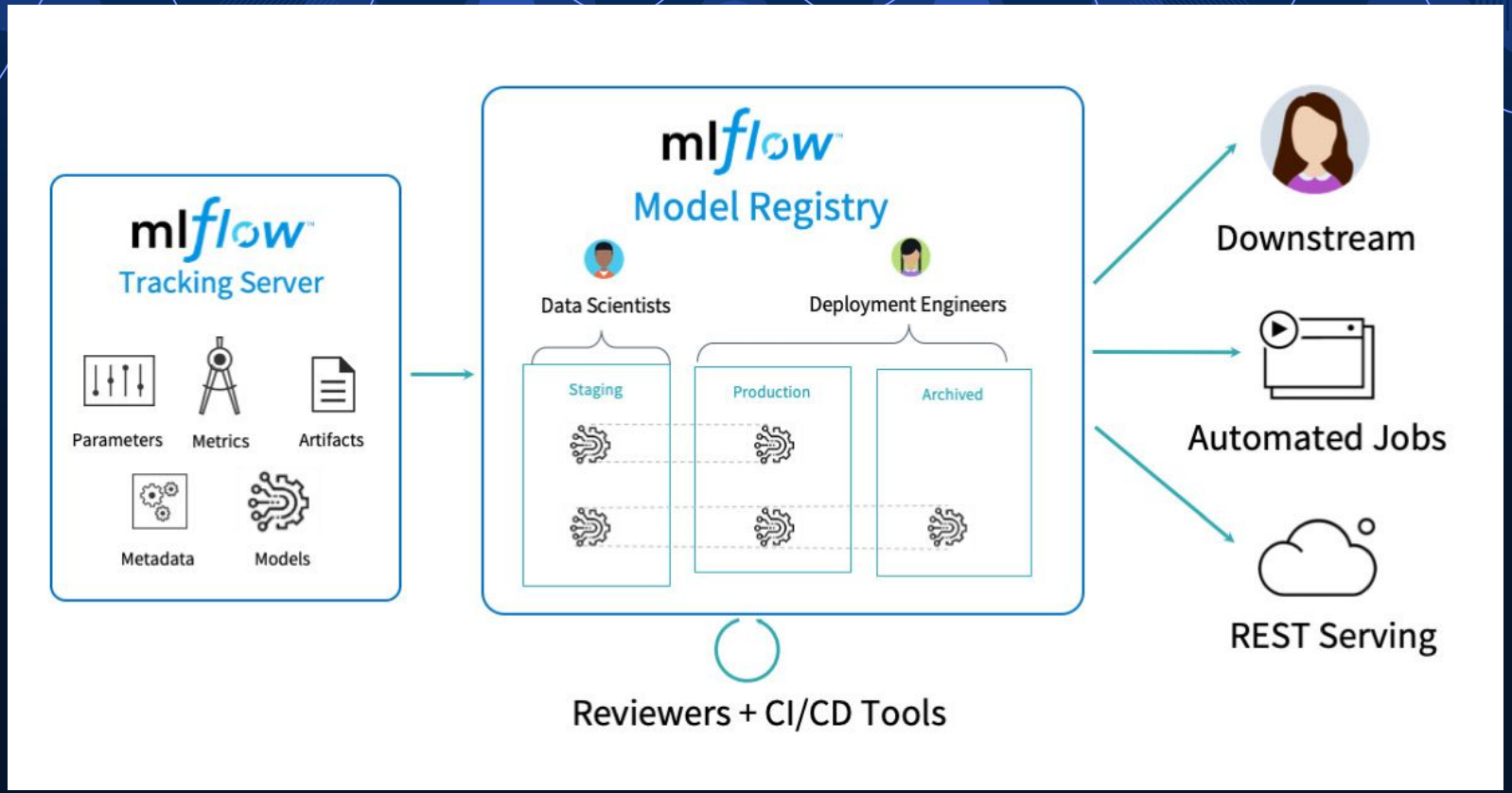
Frontend

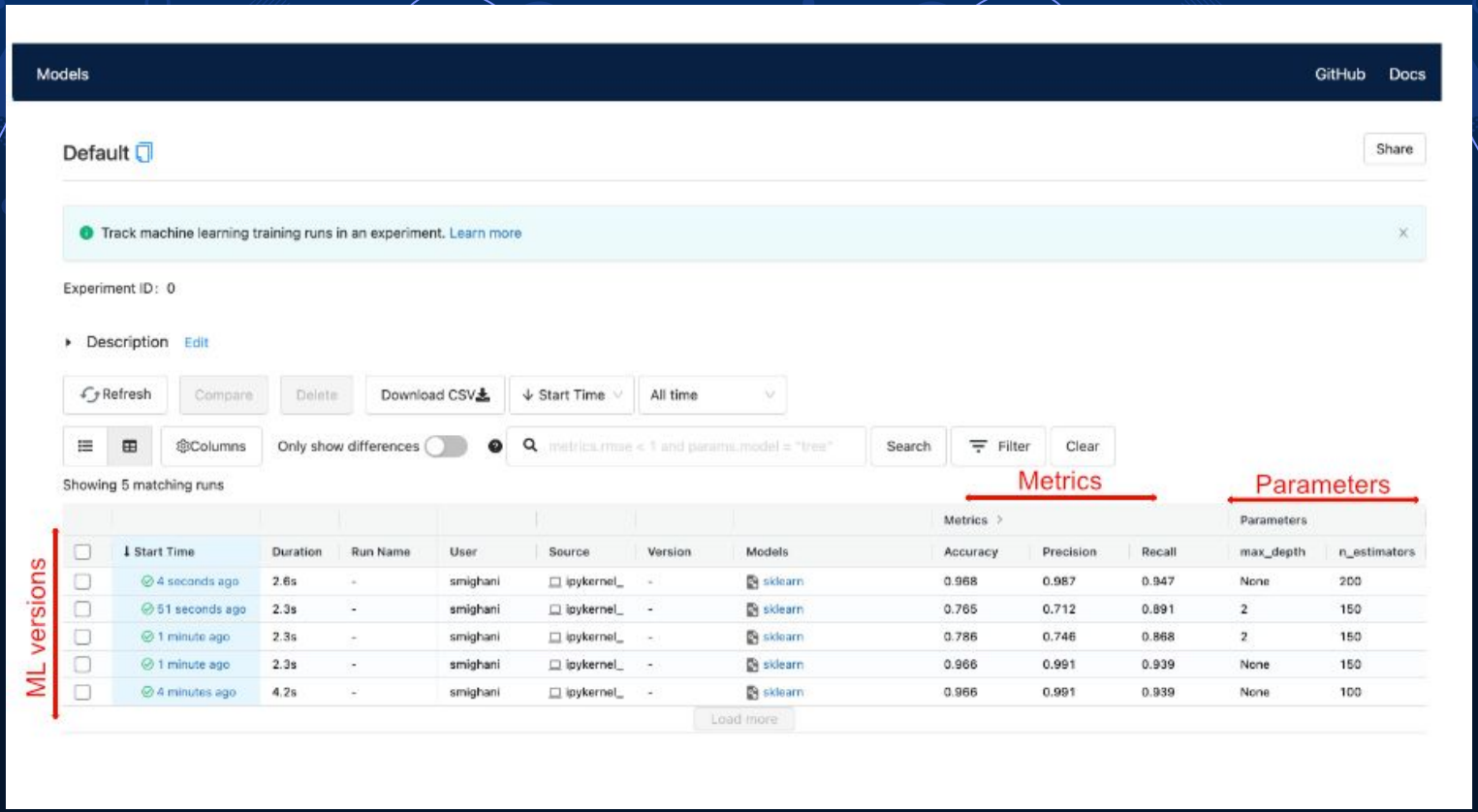
HuggingFace Spaces

Frontend + Backend

Streamlit Cloud

Level 2 - MLFlow

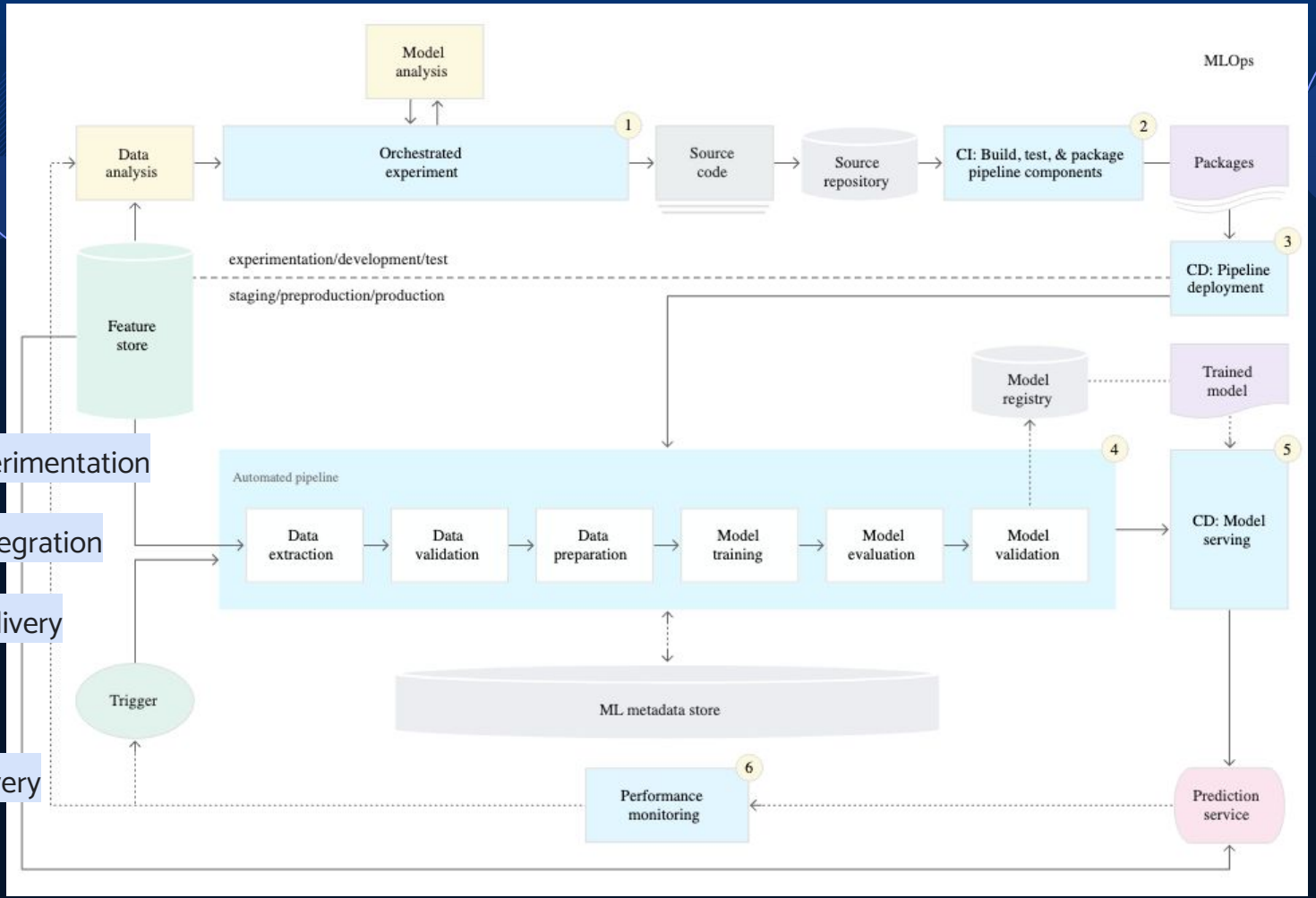


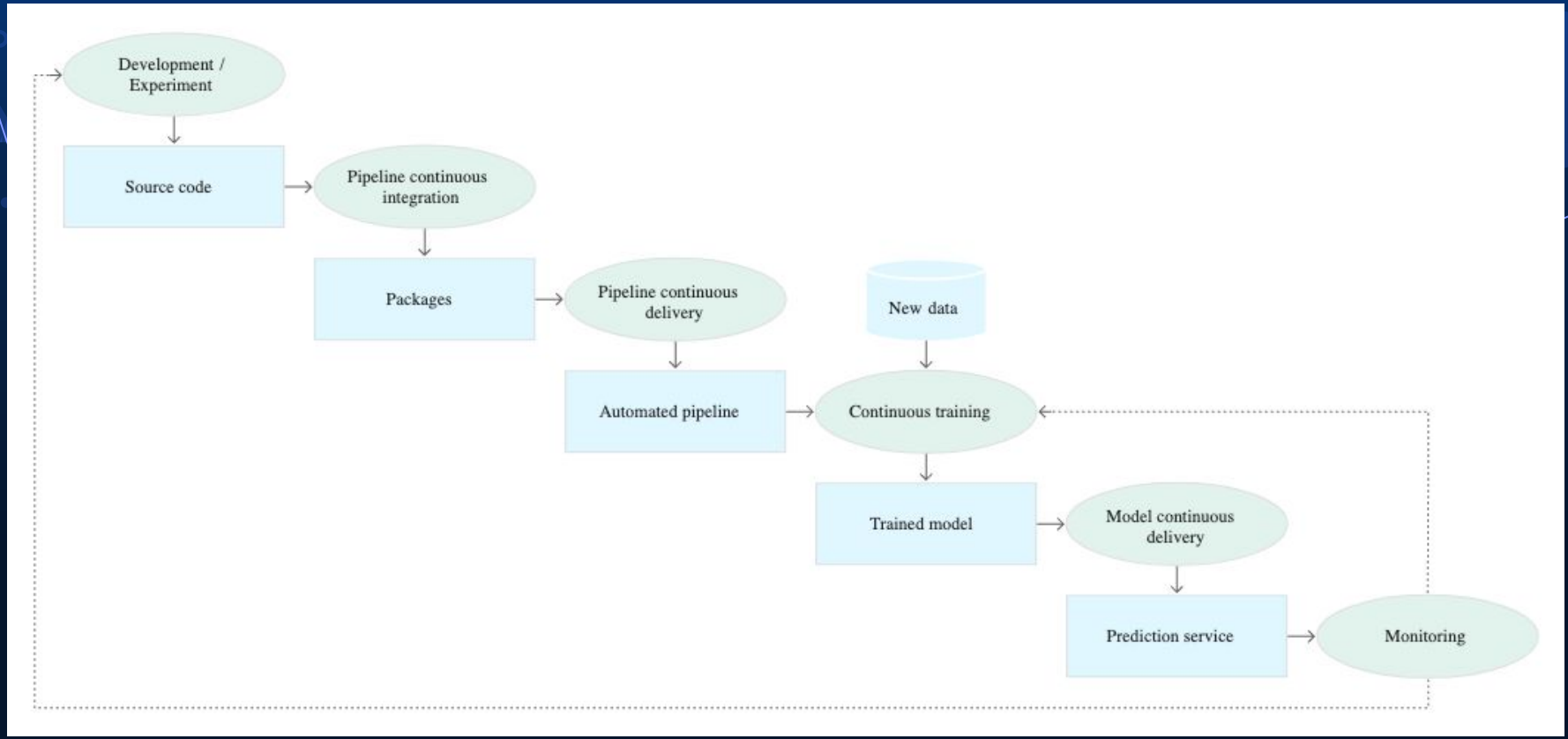


MLFlow Main Dashboard

Level 3

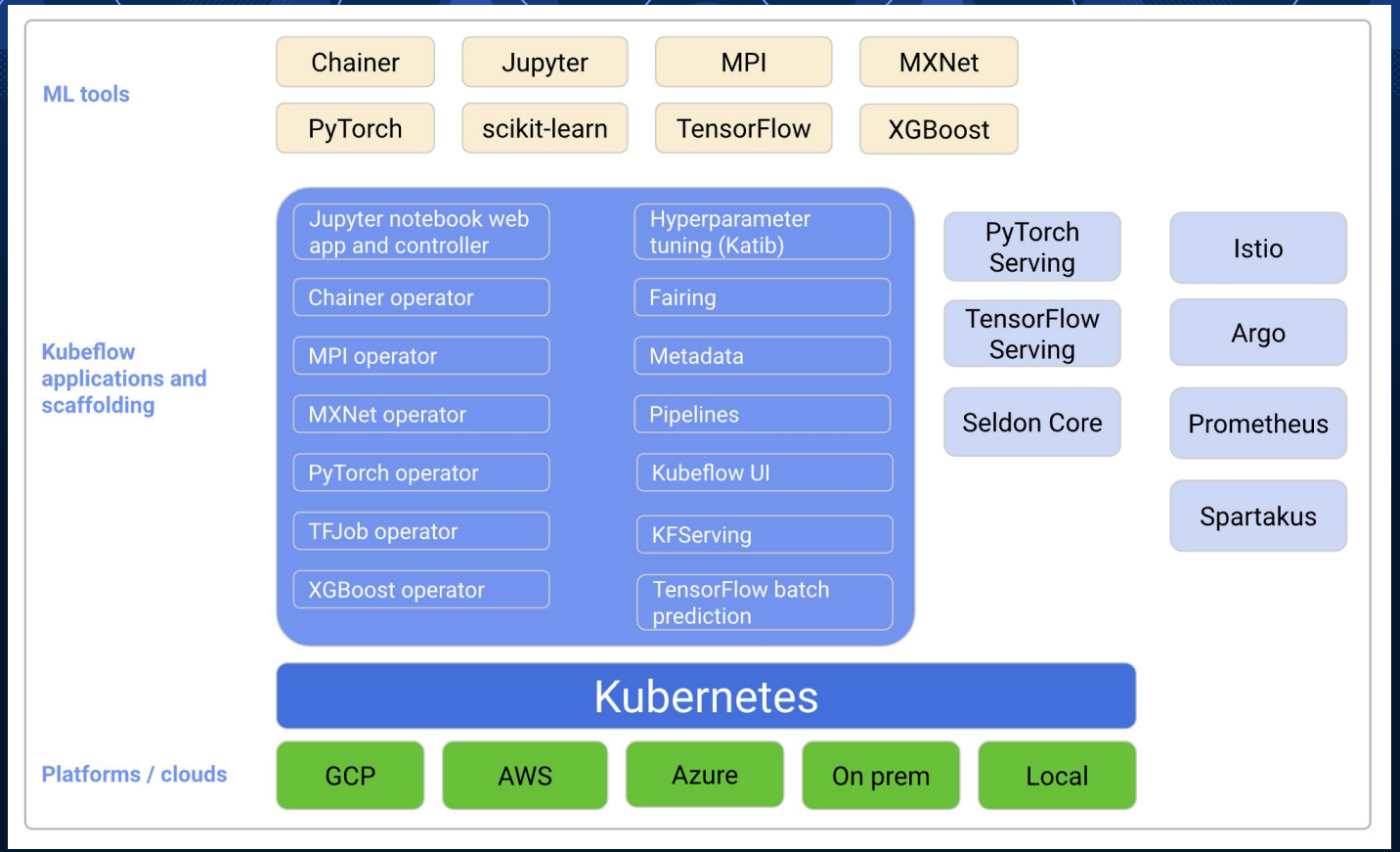
1. Development and experimentation
2. Pipeline continuous integration
3. Pipeline continuous delivery
4. Automated triggering
5. Model continuous delivery
6. Monitoring



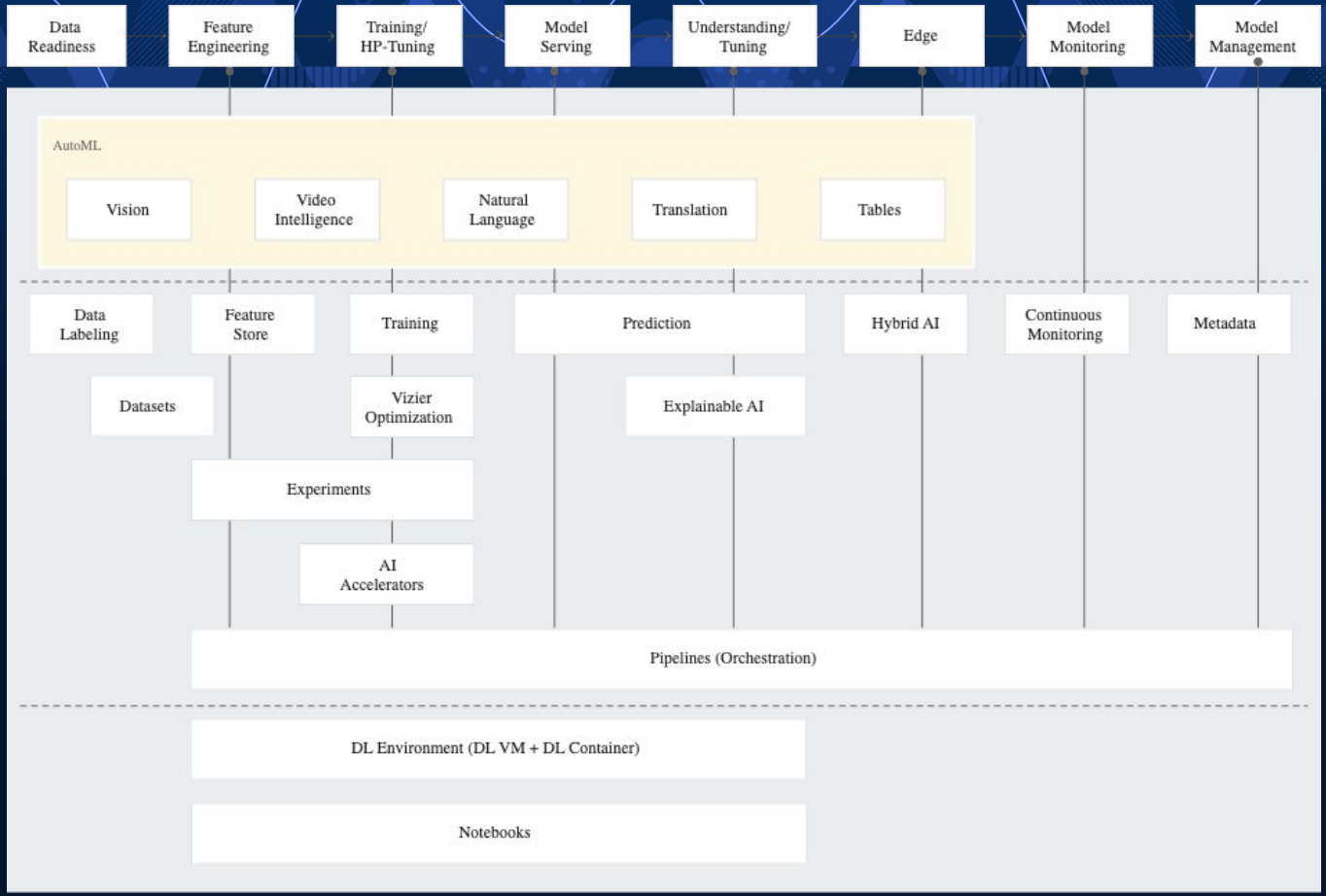


ML-Ops State Machine

Level 3 - Kubeflow



Level 3 - VertexAI (GCP)

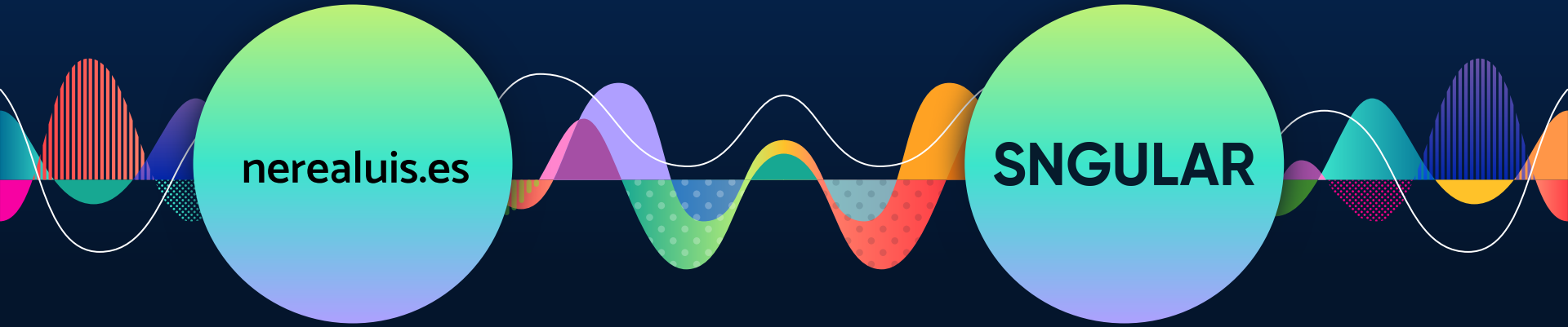


Main Conclusions

- Data Scientists and Data Engineers cannot work alone anymore. These profiles should be complemented with Backend, DevOps, QA
- ML-Ops is still a WIP field. The need to move AI models to production environments is increasing. The field will grow, lot of work to be expected here
- We can start by providing to Data Scientist simple API+Frontend to experiment with Model Serving. Then, Model Registry and Tracking Server
- Whole experience includes Continuous Delivery, Continuous Integration and Continuous Training

Thanks!
Any questions?

We are hiring!



nerea.luis@sngular.com

ai.sngular.com

Template: Slides Carnival