

We build a ML pipeline after we deploy

by Alyona Galyeva
for EuroPython 2021

Agenda

+

o

1. Introduction

2. The awesome solution:

- ML pipeline out of the box in 1 line of code
- add ML pipeline anytime even in prod env

Agenda

+

•

○

1. Introduction

2. ML Pipeline:

- why and when
- building blocks
- engineering
- debugging and monitoring
- open-source Python libraries to save your time

+

○

●

Introduction

+

o

•

ML Pipeline: why and when

Why?

reduce the cost of data science projects:

- focus on new cases/models
- prevent bugs
- audit

When?

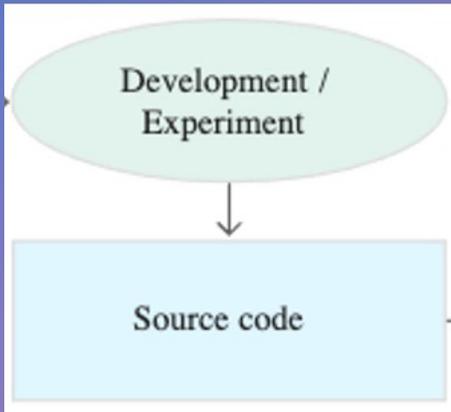
- go from PoC to MVP
- time to scale

+

o

•

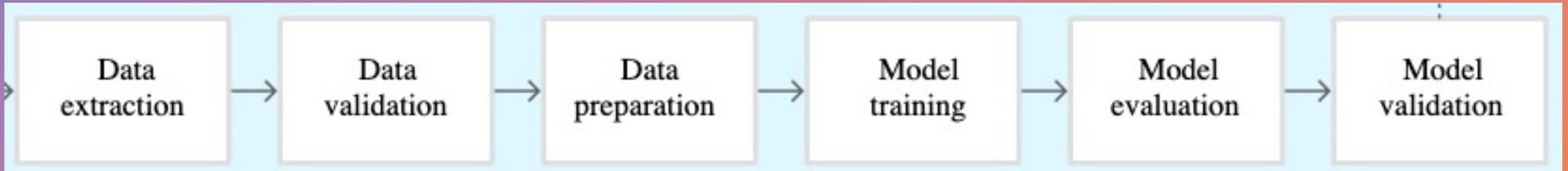
ML Pipeline: building blocks

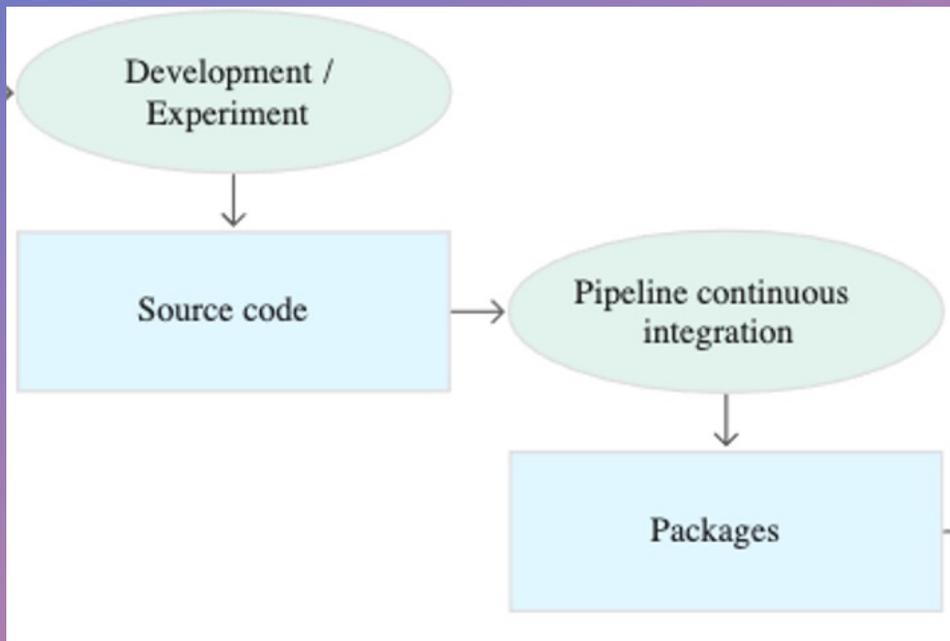


Development and experimentation

Try new ML algorithms with orchestrated experiment steps

Output: ML pipeline steps source code

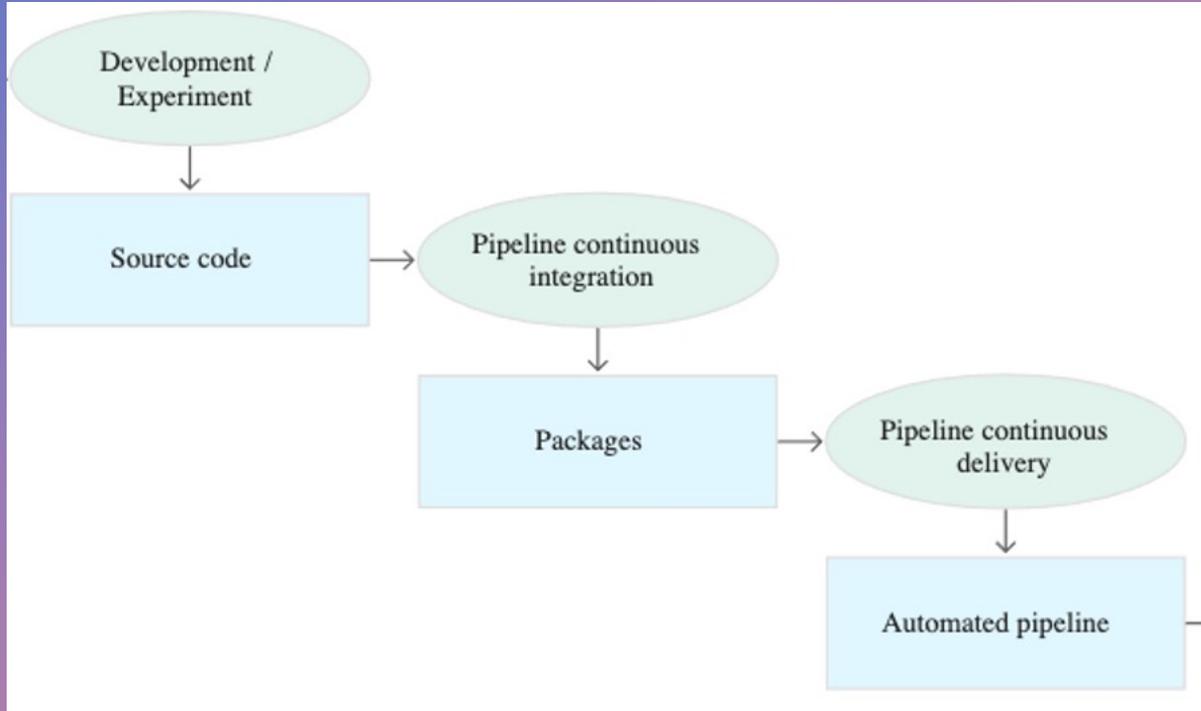




ML pipeline continuous integration

Build source code, run tests

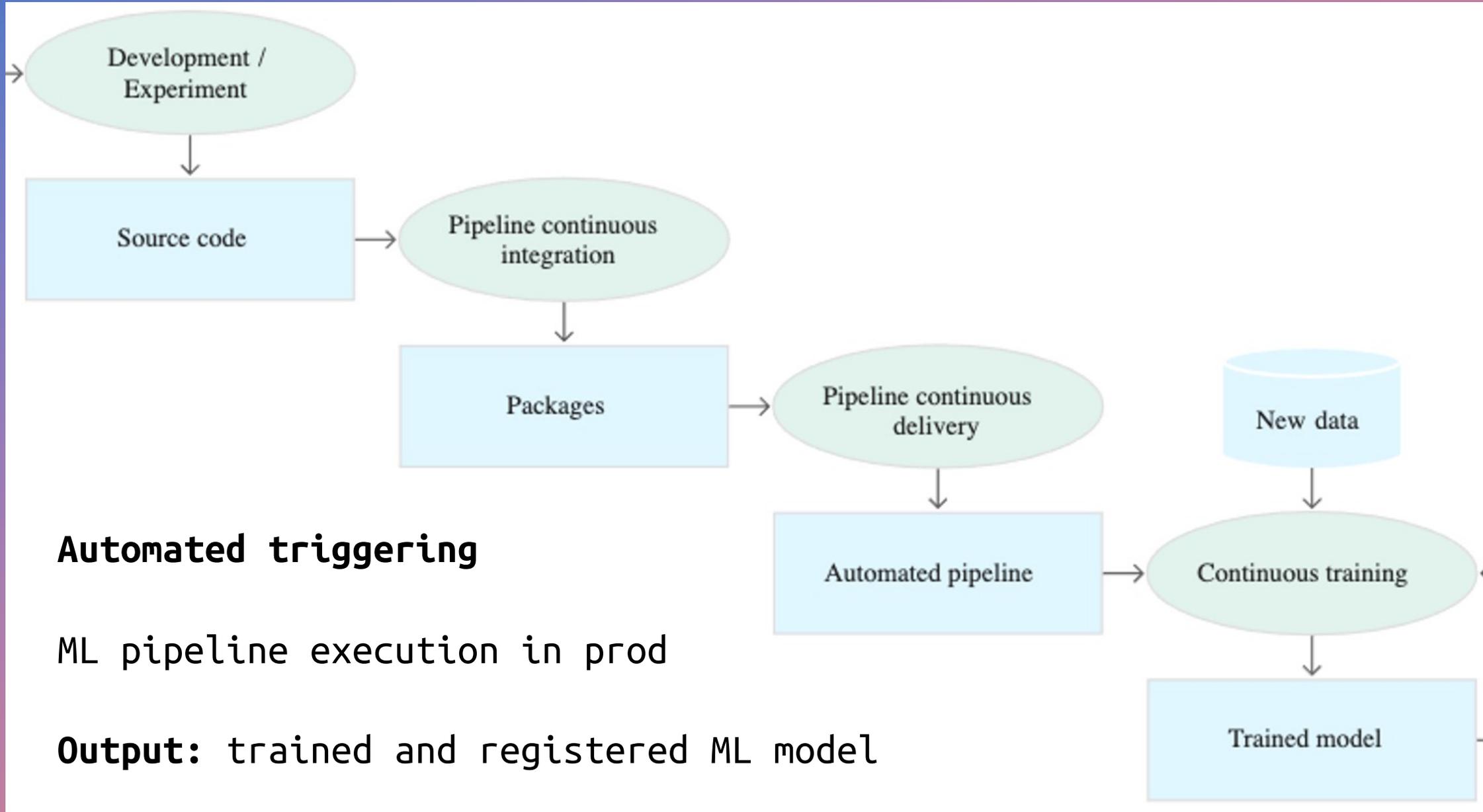
Output: ML pipeline components to deploy

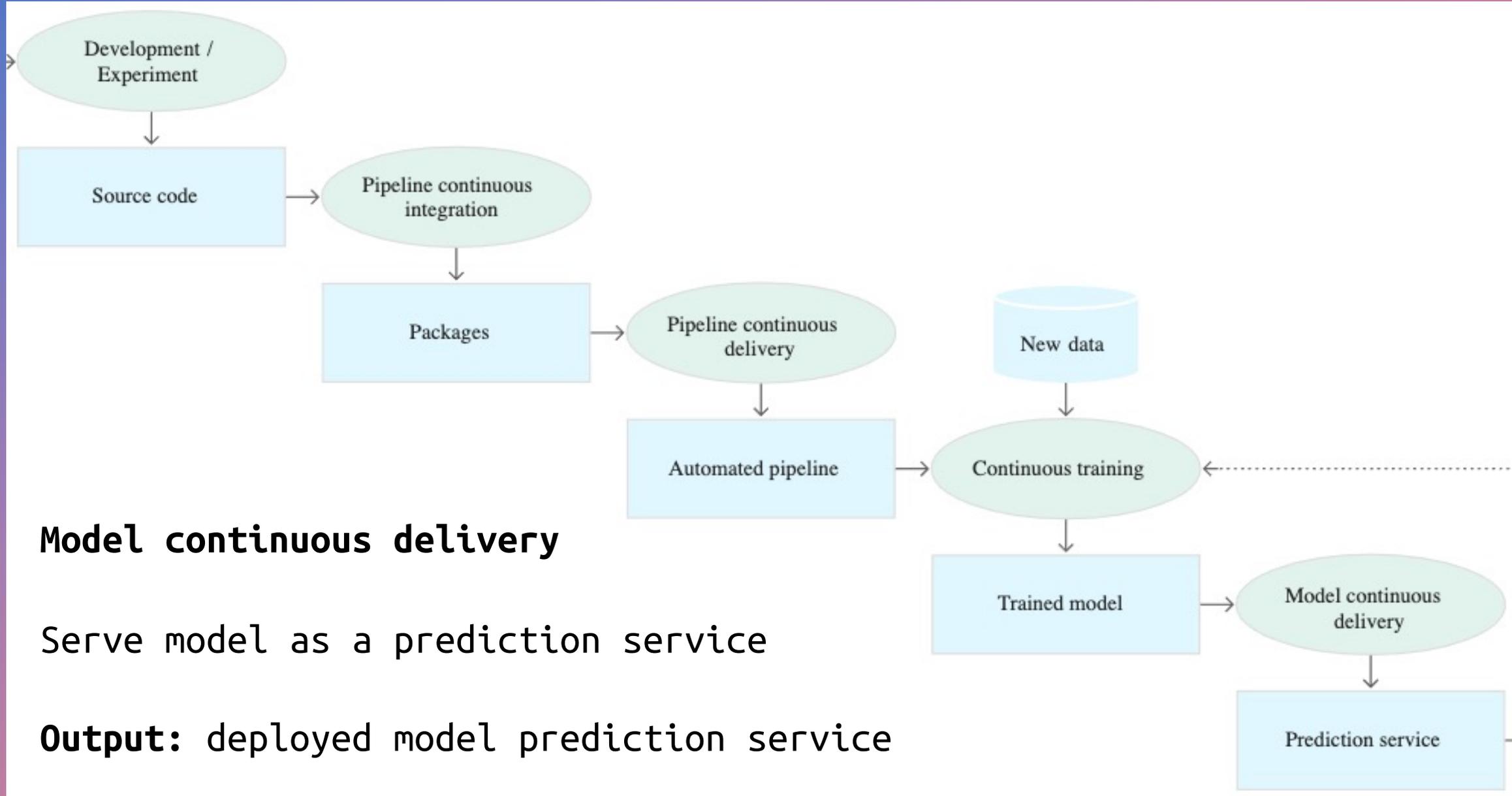


ML pipeline continuous delivery

Deploy artifacts to the target env

Output: deployed ML pipeline

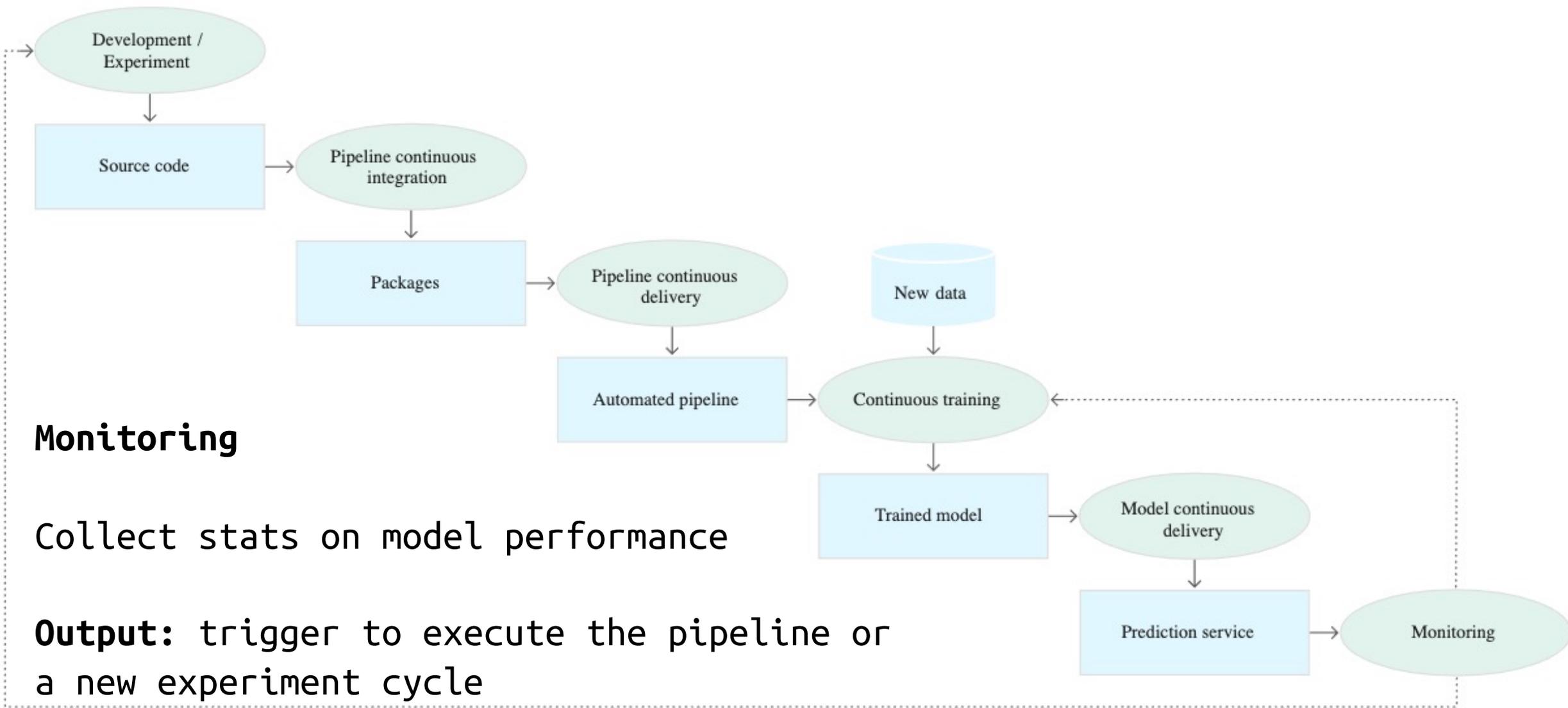




Model continuous delivery

Serve model as a prediction service

Output: deployed model prediction service



Monitoring

Collect stats on model performance

Output: trigger to execute the pipeline or a new experiment cycle

+

o

•

ML Pipeline: engineering

Engineer around failures

- input checks
- output checks
- model fallback

Engineer for performance

- scale
- caching
- feedback collection

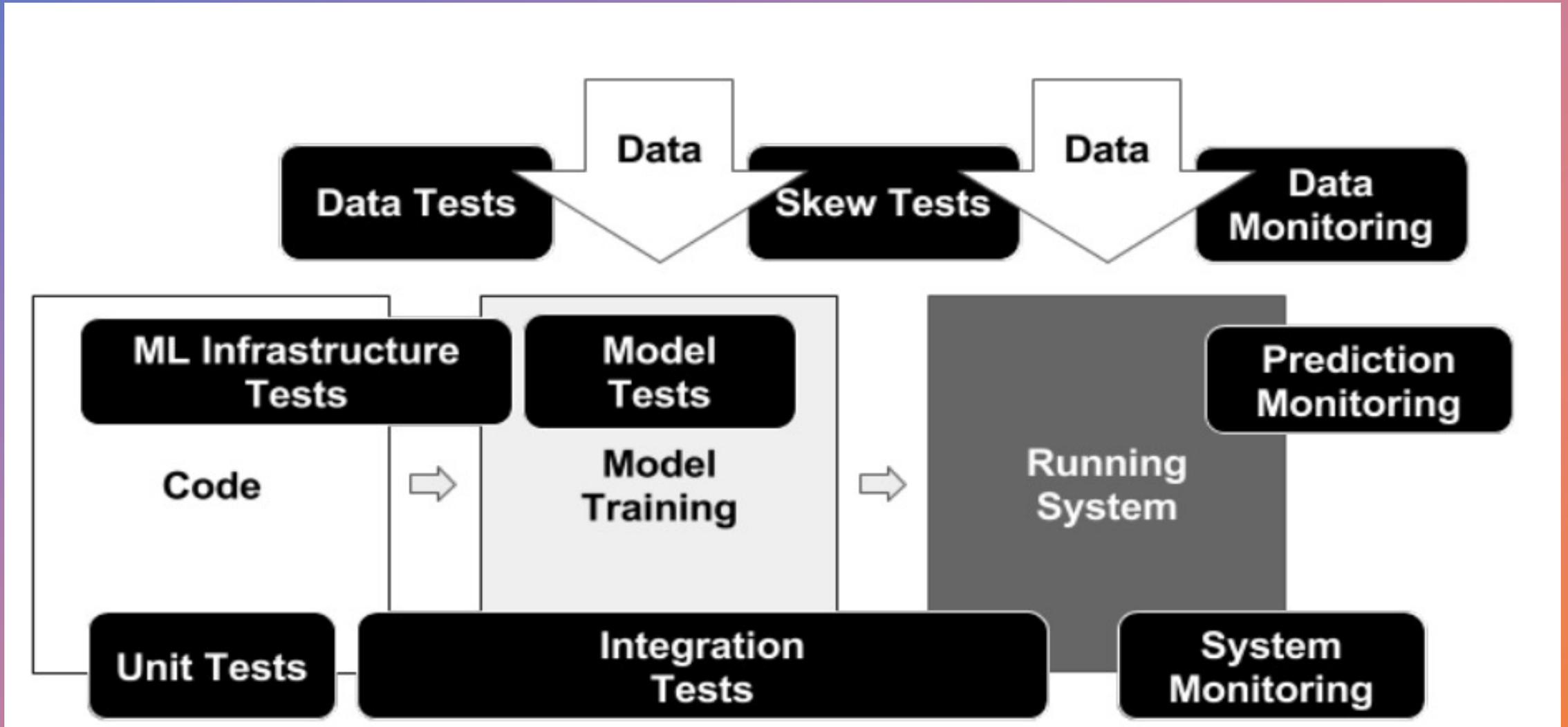
+

o

•

ML Pipeline: debugging and monitoring

ML debugging and monitoring



Top 3 debugging issues

- unpinned libraries
- data pre-processing
- scattered config for diff envs

Monitoring

- system monitoring
 - data monitoring
 - model monitoring

+

o

•

ML Pipeline: Python libraries

Delta lake, DVC - data versioning

Airflow - data processing

Great Expectations - data validation

Feast - feature store

Hyperopt, Katib - tuning

Kubeflow, Pachyderm, TensorFlow Extended - platform

LIME, SHAP, Alibi Explain - model interpretability

Mlflow - lifecycle management

Seldon Core - model serving

FastAPI - restful APIs

Pytest, locust - testing

Evidently, Alibi Detect - debugging and monitoring

Questions?