



Machine Learning Engineering Discipline in Ukraine

Yevhen Plaksa

How many disciplines are there in Data & Analytics Practice?

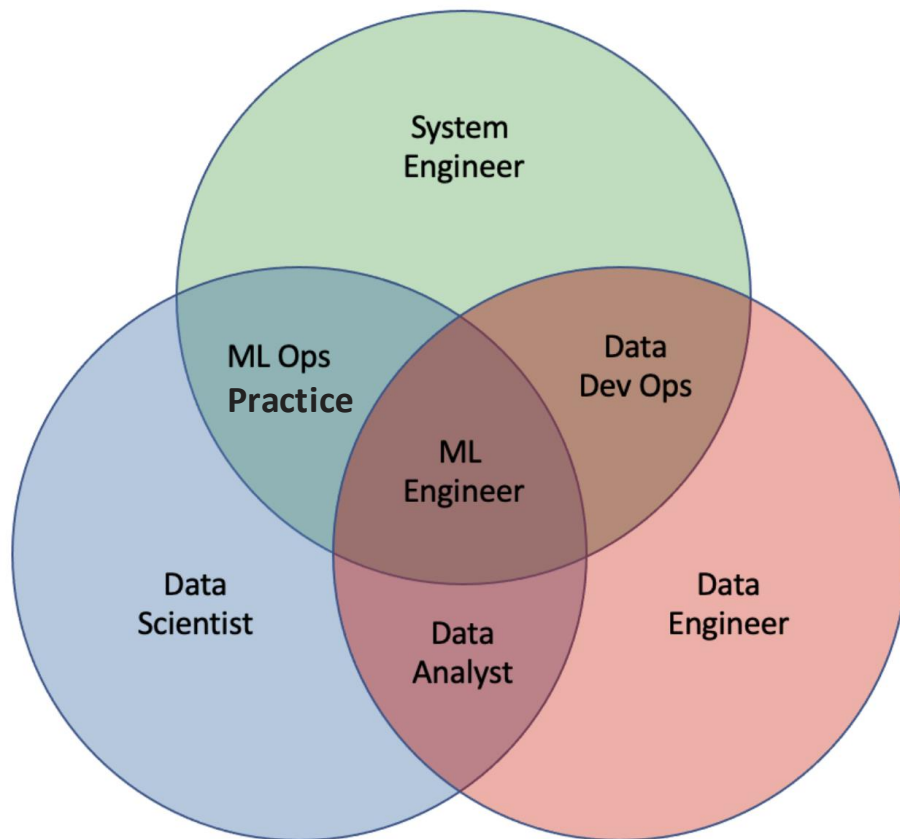
MLE discipline in EPAM locations



Discipline Attribute	Discipline Attribute Information / Links
Discipline name	Machine Learning Engineering
Relevant Primary Skill	Machine Learning Engineering
Relevant Job Function	Software Engineering
Title Lines	Software Engineer, Data Software Engineer
Discipline Placement	Part of Data Practice. See Data & Analytics Disciplines Matrix
Global Discipline Head Name	@Dmitriy Suslov
Need of Country Discipline Heads	<p>Yes. Country Discipline Heads are to be identified and assigned.</p> <ul style="list-style-type: none">Russia - @Olga PolubelovaUkraine - @Yevhen PlaksaPoland - @Radoslaw SyphenBelarus - @Kanstantsin KarneliukHungary - @Valentin Mihajlenko
Assessed Skill Potential	<ul style="list-style-type: none">The demand is growing.9 engineers have been identified who already have the necessary knowledge and skills to obtain MLE as a primary/key skill.Data Practice needs at least 100 MLE globally by the end of 2022

https://info.epam.com/topics/global/competencies-technologies/articles/2022/may/data-disciplines_en.html

What is Machine learning engineering



MLE Skills

kb MLE TI Template

<https://levelup.epam.com/titleExplorer>

Skill Category	Skill	Level	Details	Machine Learning Engineering for Production (MLE Ops)	Data Engineering Program
Data science	Basics	Basic	What is ML, cases,		
	Modeling process	Basic	CRISP-DM, train/test, etc...	v	
	Exploratory Data Analysis	Basic			
	Data cleansing	Basic		v	
	Feature generation	Basic		v	
	Basic Algorithms	Basic	SV, Regs, Clustering, Trees, etc...		
	Advanced Algorithms	Basic	NN	v	
	Quality measuring	Basic		v	
Data engineering	Optimizing a model	Basic		v	
	Basics	Intermediate	Data platform, data pipeline, data product, etc...		v
	Batch processing	Intermediate	Concept, patterns, cases, tools		v
	Streaming processing	Intermediate	Concept, patterns, cases, tools		v
	Data ingestion	Intermediate	Concept, patterns, cases, tools		v
	Data transformation	Intermediate	Concept, patterns, cases, tools		v
	Data storage	Intermediate	SQL/NoSQL, RDBMS, Hadoop, S3		v
	Orchestration	Intermediate	Airflow, etc...		v
DevOps	BigData processing	Intermediate	Spark, etc...		v
	Basics	Intermediate			
	R&D environment	Intermediate			
	Deploying	Intermediate	CI/CD, Versioning, etc...	v	
	Models as a Service	Intermediate	API	v	
	Containerisation	Intermediate	Docker, etc...	v	
	Scaling	Intermediate	Kubernetes, etc...	v	
	Monitoring and logging	Intermediate		v	
ML engineering					
	Model lifecycle	Advanced	creating, validating, testing, operating, degra	v	
	Labeling and validating	Advanced		v	
	Feature engineering	Advanced	feature store, etc...	v	
	ML pipelines	Advanced	training, inference	v	
	Inference patterns	Advanced	Batch/realtime	v	
	Model debugging	Advanced		v	
	Performance	Advanced	scheduling, latency	v	
At least one of the CLOUD Tools	Quality measuring	Advanced	A/B test, feedback loop, baseline, etc...	v	
	Deploying	Advanced	Model registry	v	
	AWS Sagemaker	Advanced			
	GCP Vertex AI	Advanced			
	Azure ML Studio	Advanced			

[Teams Group "EPAM Data Science \(& MLE\) UA Community"](#)

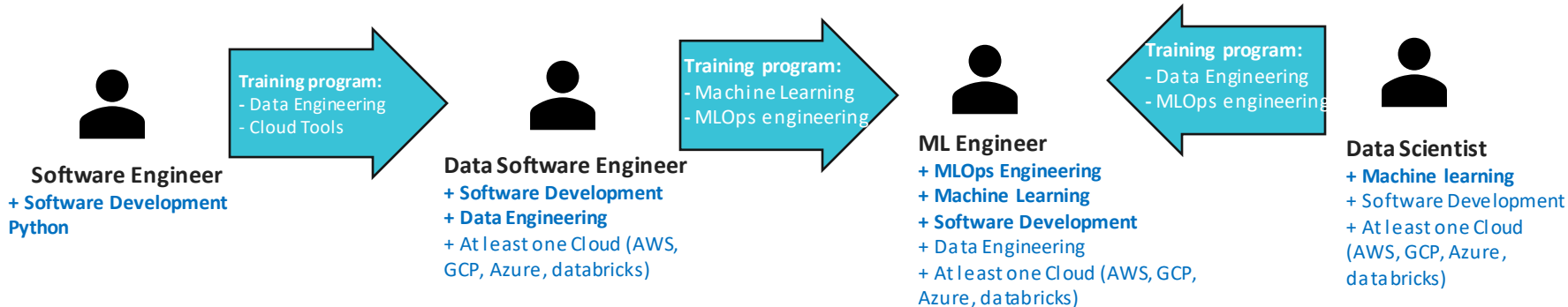
[Teams Group "Big Data Community UA"](#)

Global

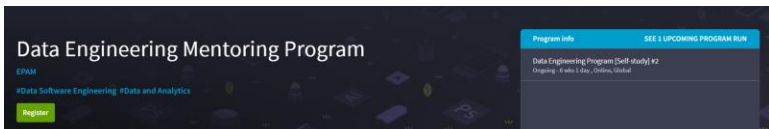
[Teams Group "EPAM Data Machine Learning Engineering"](#)

[Teams Group "EPAM Data Science"](#)

MLE training



Data Engineering Mentoring Program



Description

The Data Engineering Mentoring program covers an extended stack of modern technologies applied on the latest Data projects to satisfy the current demand for data-driven solutions and platforms. The program is enriched with additional materials and adjusted to accelerate the learning process and provide an understanding of real tasks performed on production projects.

The program has no enrollment deadlines and is delivered in a blended format. Participants will be able to study the theoretical parts online in a self-study mode and then practice processing the data set by applying various tools and technologies under mentors' supervision.

Each module consists of video presentations and additional materials to study, quizzes, and timed exams to check your knowledge and homework assignments to gain hands-on experience in Data Software Engineering.

At the end of the program, each participant will be assessed by an expert from the Data Engineering team - there is a final screening to test the knowledge gained during the program. In case of successful completion of all the modules and final screening, program participants will be able to join projects as Data Software Engineers.

Course Goal

After completion of the program attendees will be able to cover the requirements for Data Software Engineer level 2 according to the [Skill Matrix](#).



<https://kb.epam.com/display/EPMCBDDC/ML+Engineering+Courses>

Machine Learning Engineering for Production (MLOps)



Congratulations! You have completed all four courses of Machine Learning Engineering for Production (MLOps) Specialization. In this Specialization, you learned how to conceptualize and maintain integrated systems. You mastered well-established tools and methodologies to build production systems that can handle relentless evolving data and continuously run at maximum efficiency. You're now familiar with the capabilities, challenges, and consequences of machine learning engineering in production and are ready to level up your career by participating in the development of leading-edge AI technology and solving real-world problems.

MLE Transfer

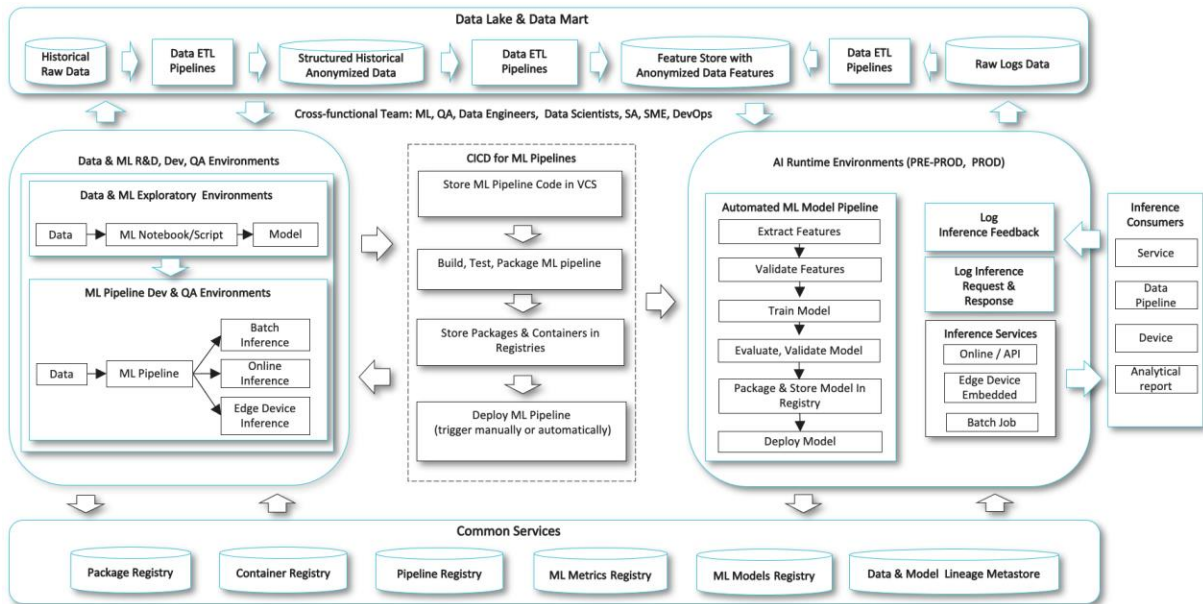
<https://kb.epam.com/display/EPMCBDC/MLE+Assessment>



SKILLS																																			
Data science								Data engineering								DevOps						ML engineering						CLOUD							
Basics	Modeling process	Exploratory Data Analysis	Data cleansing	Feature generation	Basic Algorithms	Advanced Algorithms	Quality measuring	Optimizing a model	Basics	Batch processing	Streaming processing	Data ingestion	Data transformation	Data storage	Orchestration	Big Data processing	Basics	R&D environment	Deploying	Models as a Service	Containerisation	Scaling	Monitoring and logging	Model lifecycle	Labeling and validating data	Feature engineering	ML pipelines	Inference patterns	Model debugging	Performance	Quality measuring	Deploying	AWS Sagemaker	GCP Vertex AI	Azure ML Studio
2	1	1	2	2	2	1	1	2	3	3	3	3	3	3	2	3	2	1	2	1	3	2	2	2	1	2	2	1	2	2	2	1	1	1	1

EPAM MLOps Platform Concept

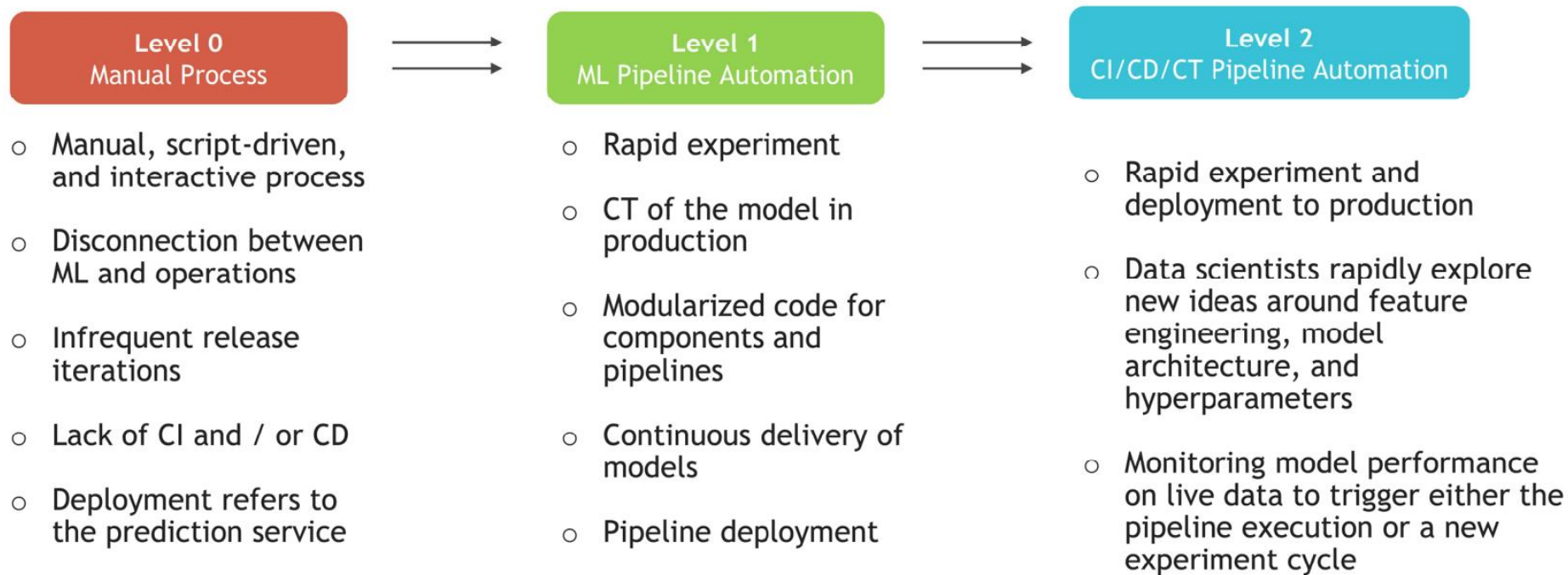
A conceptual view on ML/AI operations platform addresses challenges related to ML models operations in production environment.



Key Best Practices

- Data Mart with high quality curated data sets
- Automated data ingestion and ETL pipelines
- Isolated R&D, QA, PROD environments
- ML model learns from a feature sets located in curated feature store
- ML experiments tracking
- ML pipeline and ML model registries in addition to traditional code, containers, and package registries
- Automated ML pipelines
- Continuous Integration and Continuous Delivery (CI/CD) pipelines
- Quality gates embedded into ML and CI/CD pipelines
- Feedback loop automation
- Data drift detection
- Model interpretability
- Batch, online, and embedded in edge devices inference capabilities

MLOps Maturity Model



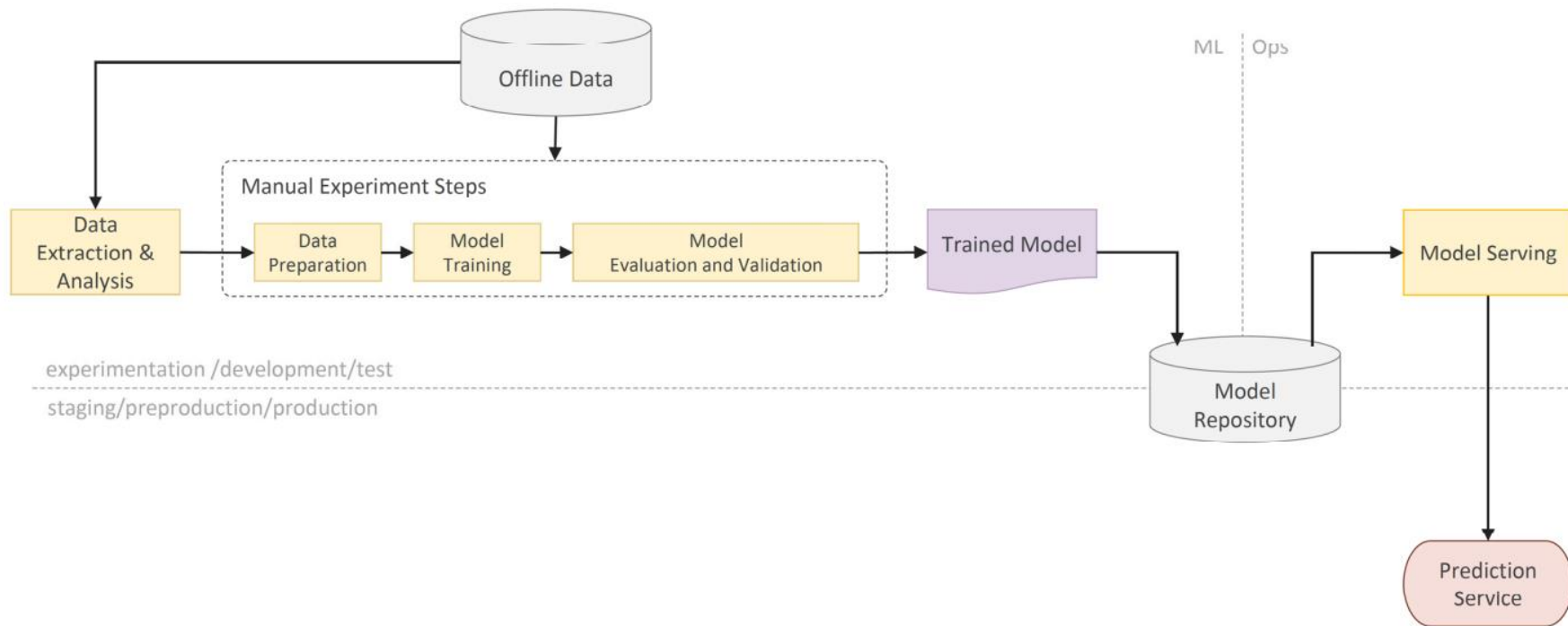
[EPAM_MLOps_Solution_v0.9.5.pdf](https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning)

<https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning>

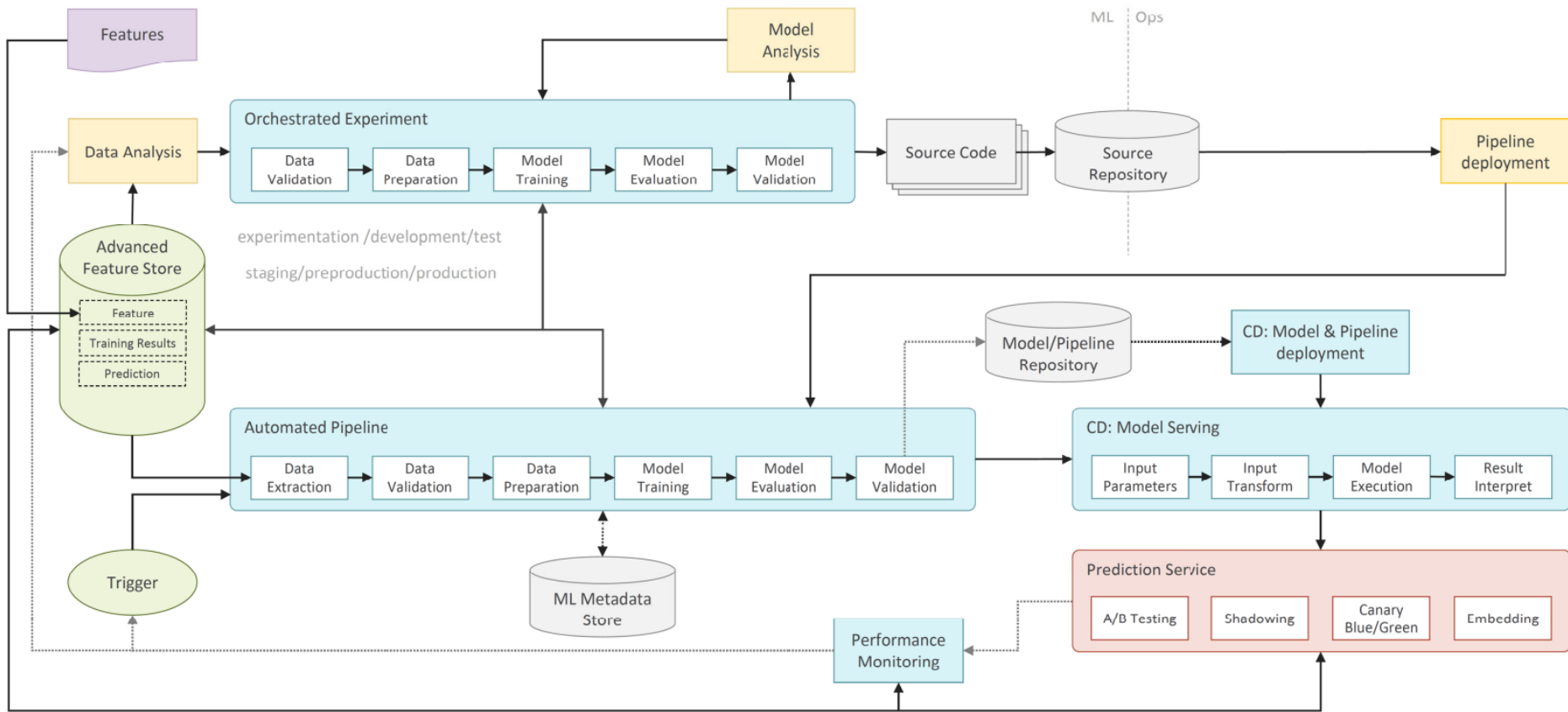
<https://docs.microsoft.com/en-us/azure/architecture/example-scenario/mlops/mlops-maturity-model>

<https://aws.amazon.com/ru/blogs/apn/taming-machine-learning-on-aws-with-mlops-a-reference-architecture/>

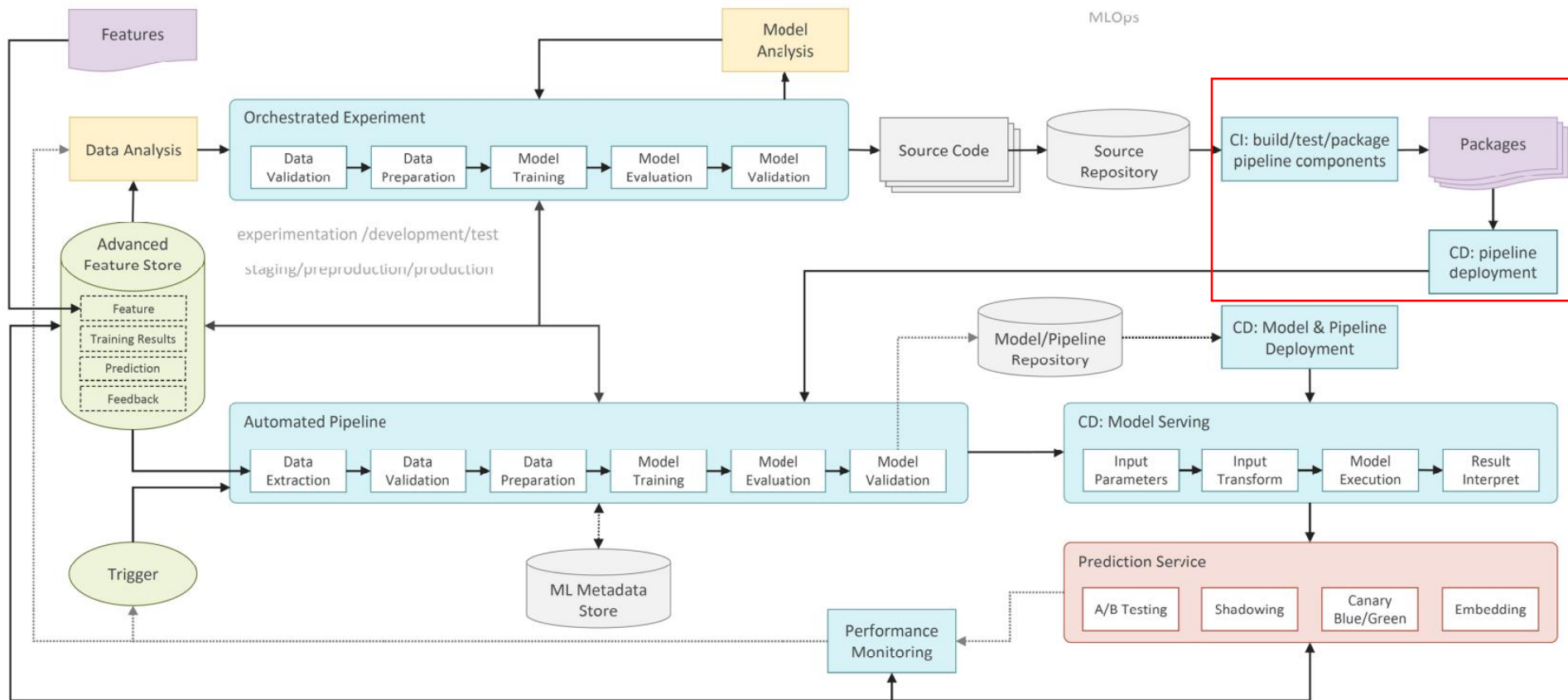
MLOps Maturity Model: Level0- Manual Process



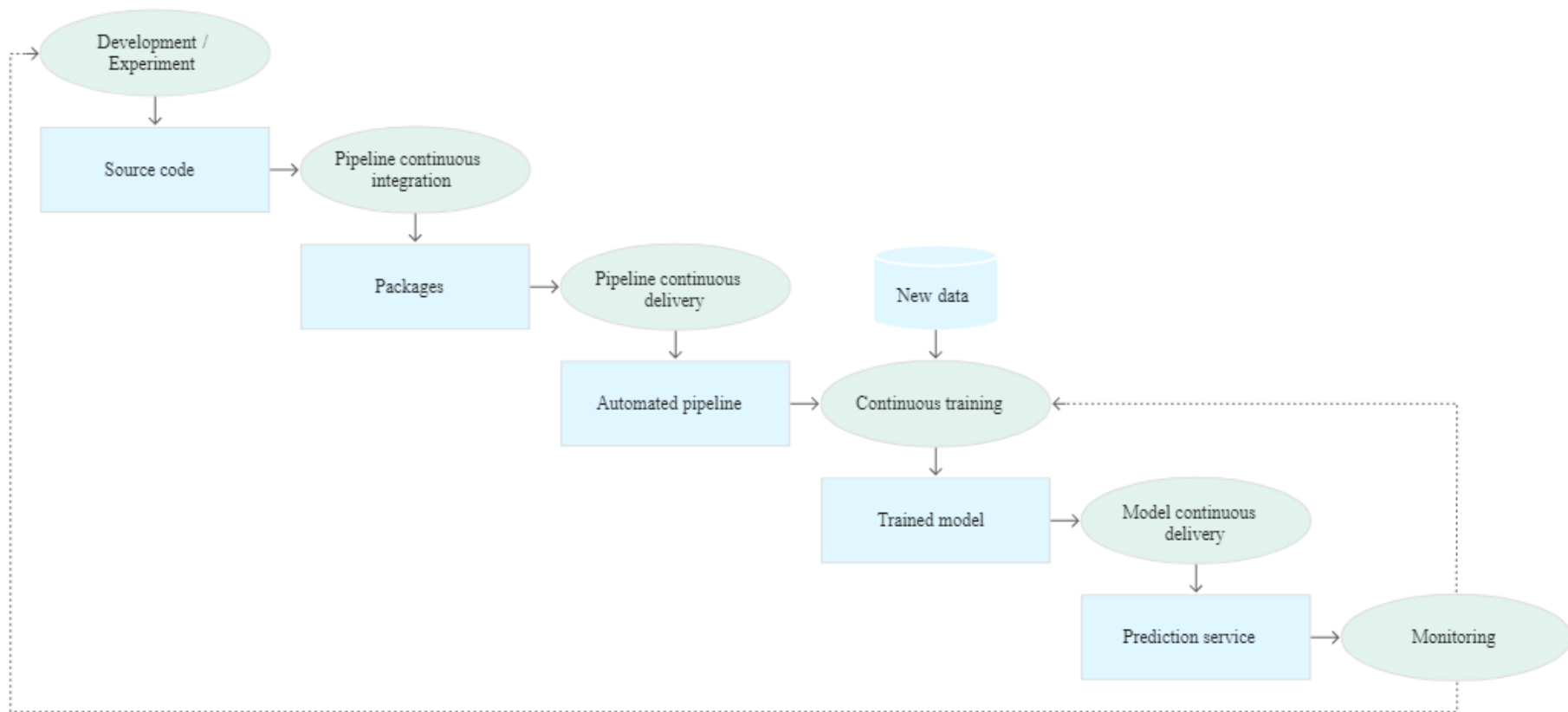
MLOps Maturity Model: Level1- ML Pipeline Automation



MLOps Maturity Model: Level2- CI/CD/CT Pipeline Automation



Putting all Together



ML QA Pyramid



Data

Schema
Sampling over Time
Volume



Model

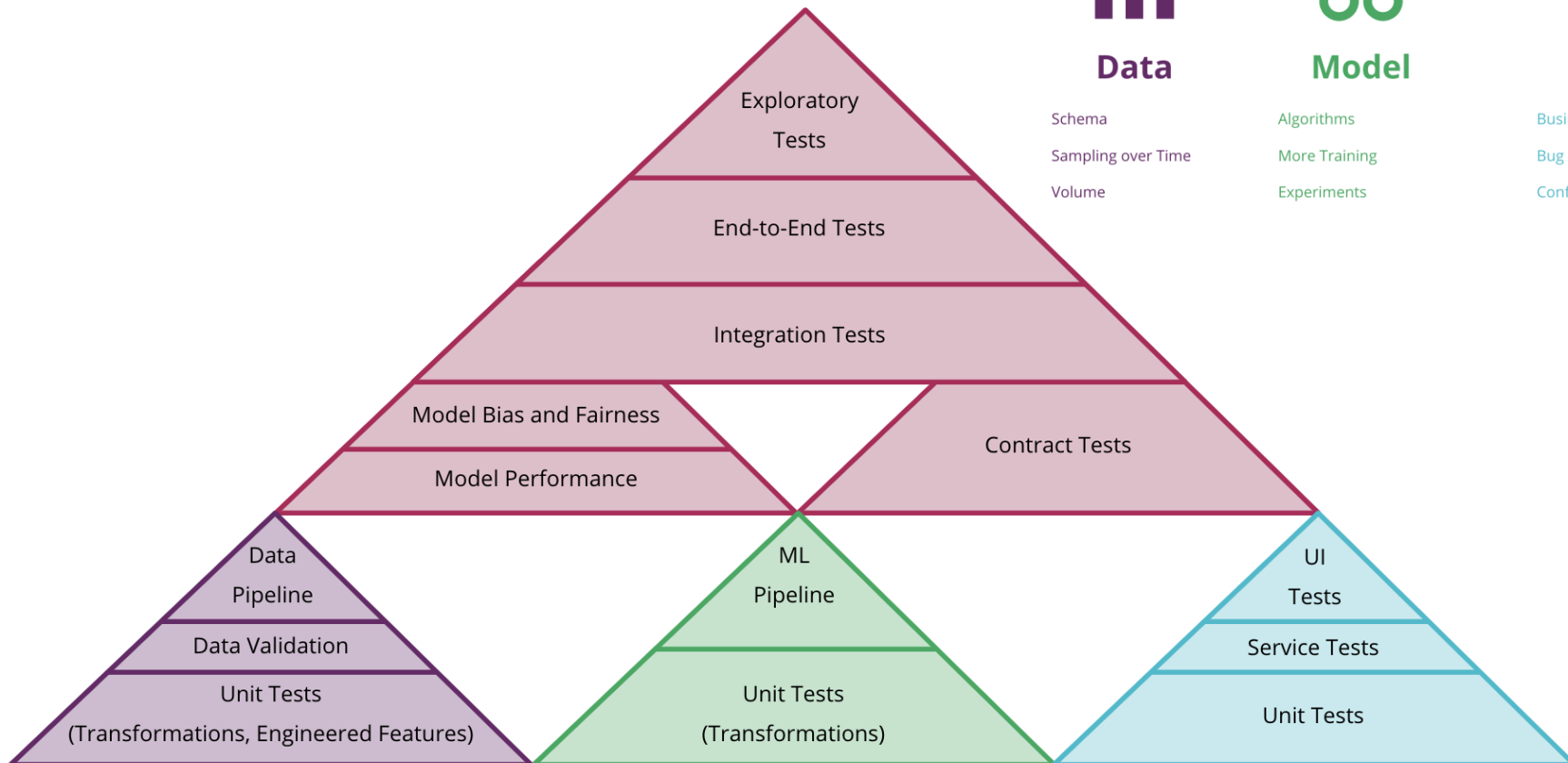
Algorithms
More Training
Experiments

+



Code

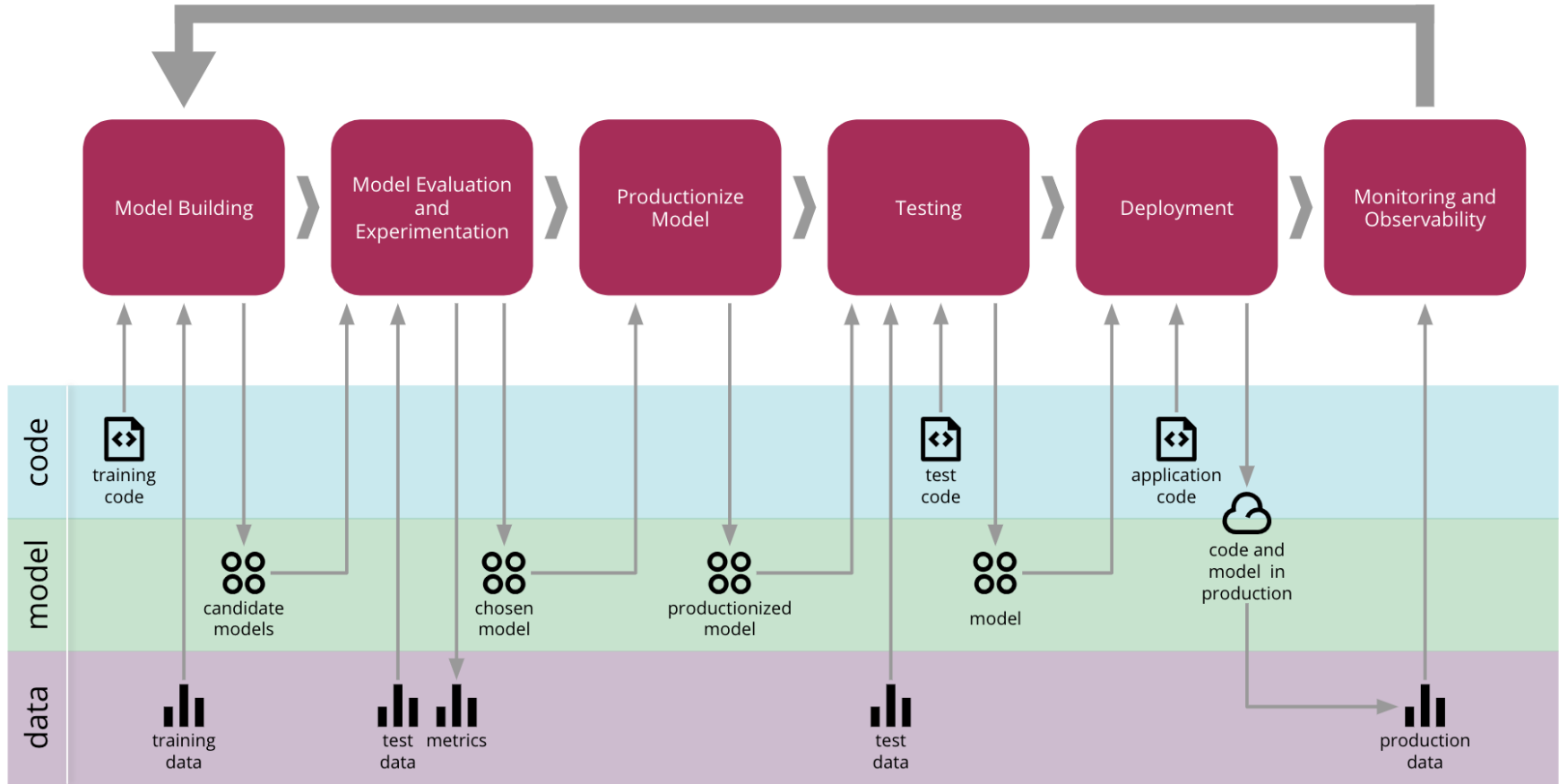
Business Needs
Bug Fixes
Configuration



<https://medium.com/mlops-community/test-driven-development-in-mlops-part-1-8894575f4dec>

<https://martinfowler.com/articles/cd4ml.html>

The End-to-End CD4ML Process



<https://martinfowler.com/articles/cd4ml.html>

