# Machine Learning Operations
# MLOps Excite

**Team MLOps**

Fraunhofer IAIS

# MLOps Excite
## Agenda

1. **Fraunhofer IAIS and Uni Bonn**
2. **Team MLOps**
3. **Machine Learning Operations (MLOps)**
   - **Introduction**
   - **Organization**
   - **The MLOps Cycle**
   - **ML Pipelines and Automation**
4. **Become a student assistant at IAIS-MLOps**

# MLOps Excite
# Fraunhofer IAIS and Uni Bonn

# Fraunhofer IAIS

## Strong Publication record

### Various influencial IAIS publications

- **Conference Papers**
  - E.g., in informed ML 2021 Informed Machine Learning Taxonomy
  - 2022: over 65 conference publications
- **Studies**
  - E.g., AI-supported design of experiments in research and development
- **White Papers**
  - Introducing a team's products
  - E.g., Efficient fraud detection: discover new potentials

# Fraunhofer IAIS

## Data Science Training Courses

### Wide range of certified training courses

- Industrypartner train data science competences
- **5 day certified Data Science Training**
- **5 day certified Data Analytics Training**
- **5 day certified Data Management Training**
- **…**
- **1 day compact Machine Learning Entry**
- **1 day compact Computer Vision Entry**
- **…**

# MLOps Excite
# Team MLOps

# MLOps Training Course
Presented by the MLOps Team

## MLOps Training Courses

- Own training courses about MLOps with a lot Hands On Samples
  - **5 day certified MLOps Training**
  - **1 day MLOps compact Training**

- Ongoing slides give an abbreviation of those courses

# MLOps Excite
# Machine Learning Operations (MLOps)

# MLOps Excite

## Agenda

1. Introduction
   - History
   - ML Context
   - Our approach
2. Organization
   - Project and product
   - Roles
   - Communication
3. The MLOps Cycle
   - Phases
   - Components and tools
4. ML Pipelines and Automation
   - Automation levels
   - ML Pipelines

Fraunhofer IAIS

# MLOps Excite Introduction

# MLOps Excite

## Introduction: History

| 1980s | 1990s | 2000s | 2010s | 2020s |
| --- | --- | --- | --- | --- |

**Consideration of the user**

- Design Thinking
- Prototyping

**Integration of the user**

- Iterations and Releases
- Scrum
- CRISP-DM Model

**DevOps**

- CRISP-DM-Modell
- Scrum, Kanban
- Hadoop, Big Data
- CI/CD

**ML Projects**

- Data Science
- Cloud
- Model optimization in operation

**MLOps**

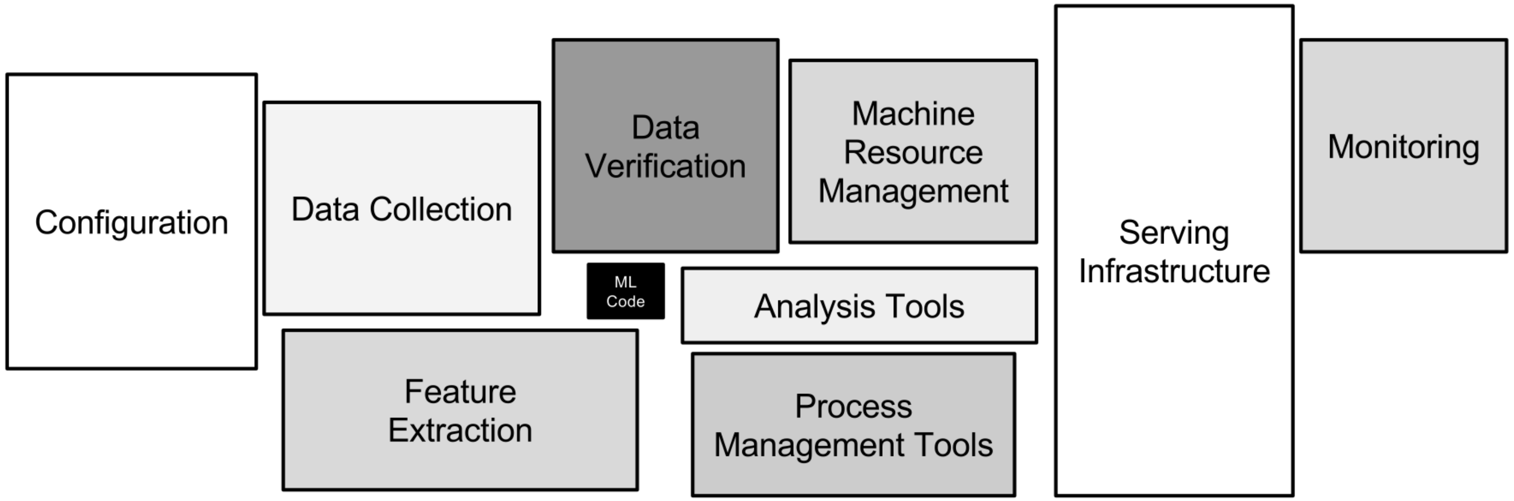
- Data Management
- Monitoring
- Drifts
- Edge

- *Drift Detection*
- *Continual Learning*
- *Auto Labeling*
- *Auto Test Data*
- *Explainability*

# MLOps Excite

## Introduction: ML Context




Extracted from: Hidden Technical Debt in Machine Learning Systems
(https://proceedings.neurips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html)

### ML Context

- Only a small part of the real ML systems consists of the ML code
- It is represented by the black box in the middle
- The required surrounding infrastructure is extensive and complex

# MLOps Excite

## Introduction: ML Context

### Machine Learning in practice

- Digitalization promotes the use of ML as a supporting element in the company
- Software has already been developed and used productively for decades
  - Established software development processes
  - State of the art: DevOps
- ML, however, harbours a new paradigm of software development
  - Data-driven learning
  - Integration of expert knowledge (Informed ML)
- CRISP-DM as a standardized procedure for Data Science projects
  - MLOps never stands alone!
  - There must always be an embedding in relevant data science projects.


Image rights are held by: Dennis Wegener, Fraunhofer IAIS

# MLOps Excite

## Introduction: ML Context

### CRISP-DM ends in Deployment

- Model application can be done in a different system/language than model creation
- Model application can be under "real-time" conditions (24/7 operation, low latency)
- Model application may have additional non-functional requirements (robustness, scalability)

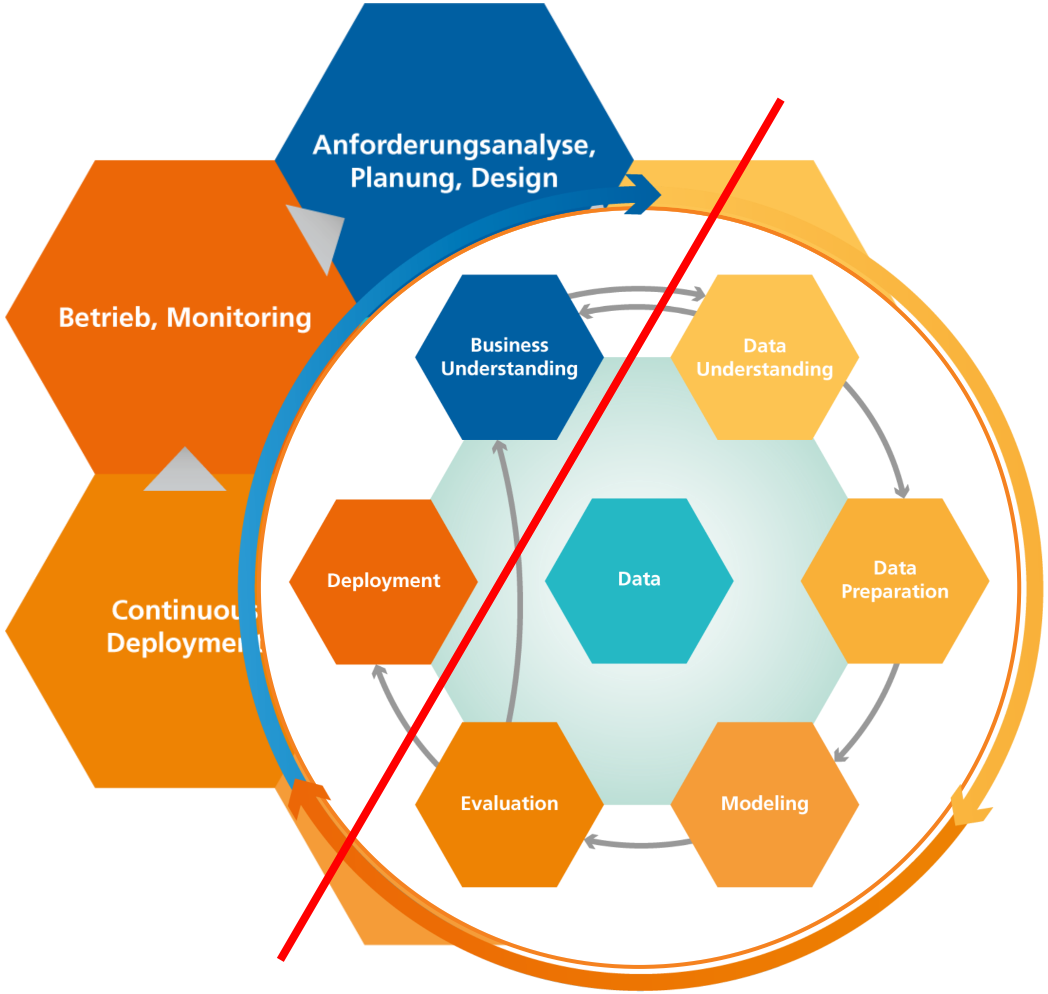



Image rights are held by: Dennis Wegener, Fraunhofer IAIS

# MLOps Excite

## Introduction: Our Approach

### Professional Integration

- Establishment of end-to-end management processes
- Agile project structure with strong integration of the business side

### Technical Integration

- Use of ML-specific tools for the complete ML life cycle
- Experience from data science projects on technical feasibility

### Scalability

- Virtualization as a core concept
- Technical architecture matching the business requirements

# MLOps Excite

## Introduction: Our Approach

### Goal

- Implement & automate the phase from "exploration" to "operation“ with tool support

### Reasons

- Combine development and operation
- Tool support & monitoring in all phases of ML system development
- Apply best practices (CI/CD) to ML systems
- React faster and more efficiently to changes in data or business context
- Establish traceability and enable testing
- Higher degree of automation enables focus of resources

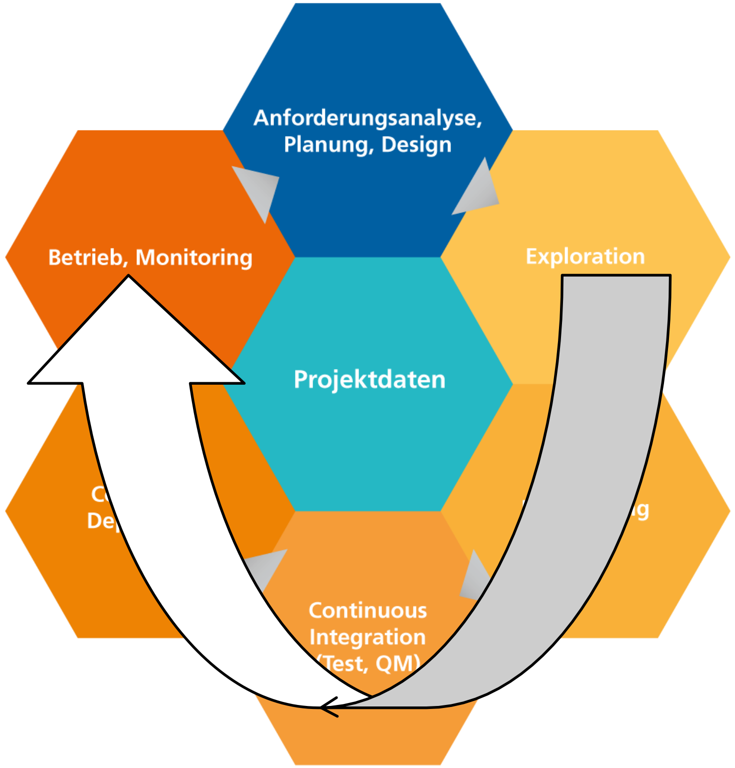



Image rights are held by: Dennis Wegener, Fraunhofer IAIS

**From "Providing Predictions" to "Models & Containers" to "Providing an Automated Pipeline"**

# MLOps Excite Organization

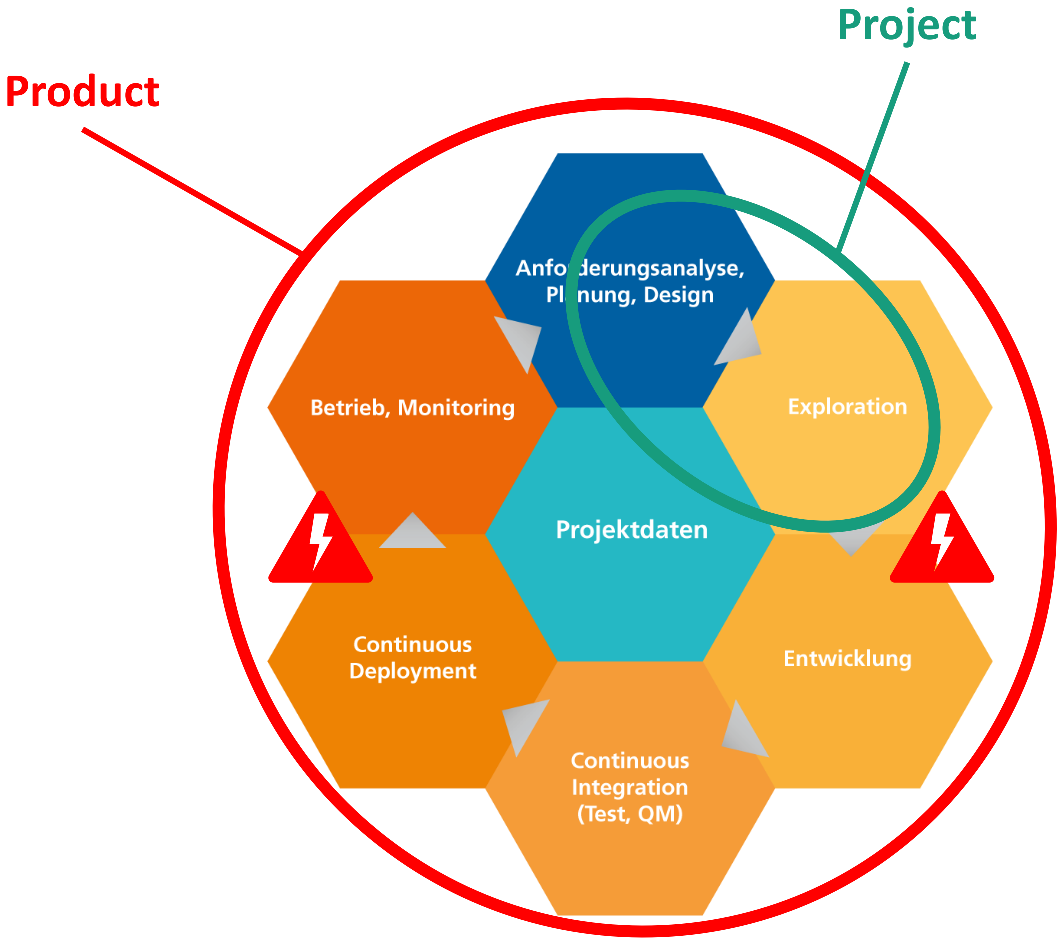
  

# MLOps Excite

## Organization: Project and Product

### MLOps is a technique to support Data Science projects and products

- Strongly dependent on the maturity of the data science projects/products
- Two elementary "breaking points" in Data Science projects
  - Transition between “exploration” and “development”
  - Automated deployment under QM aspects
- Internalize the difference between project and product!



# MLOps Excite

## Organization: Roles

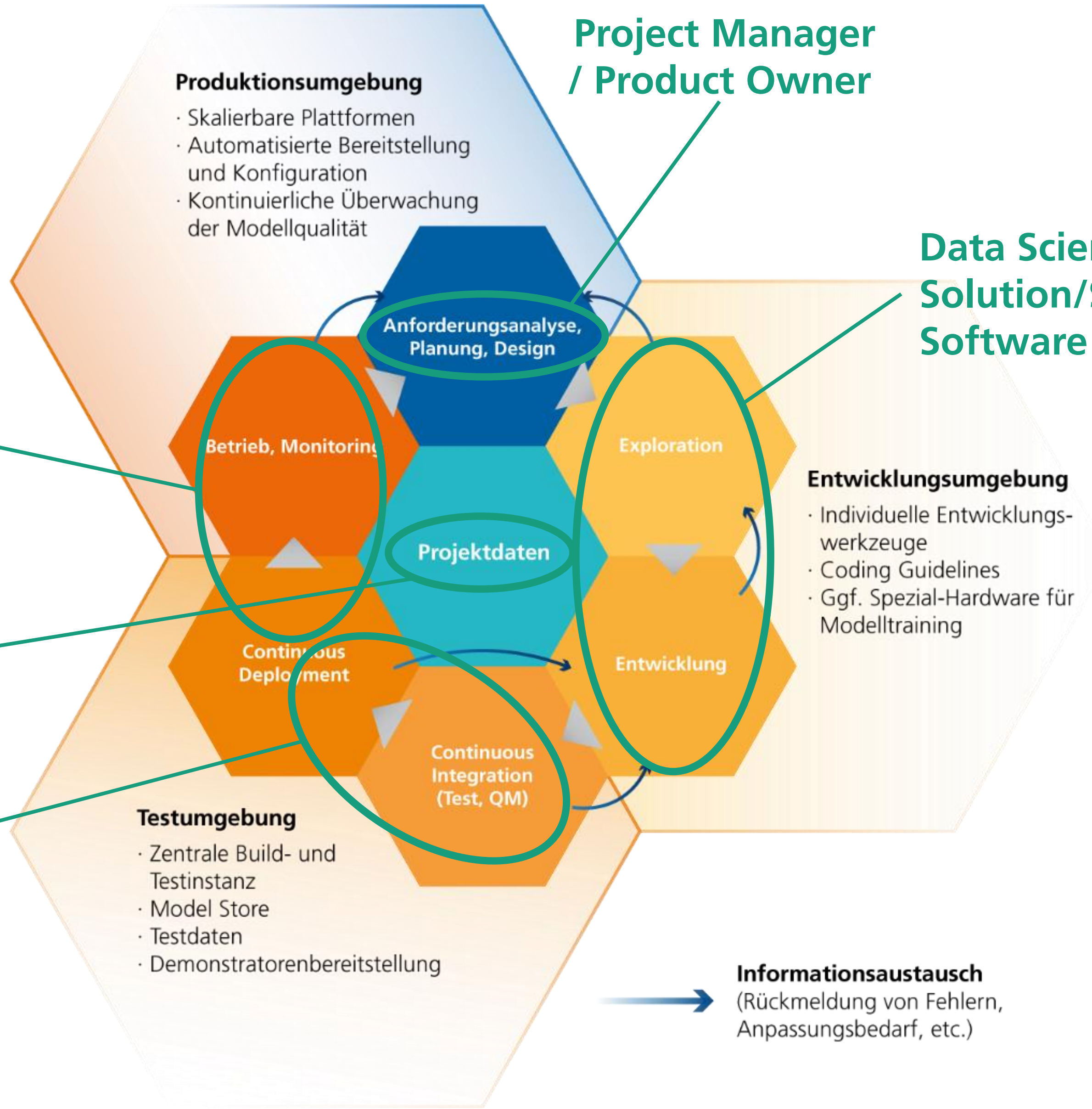
Project Manager / Product Owner

Data Scientist
Solution/Software Architect
Software Developer

Service/Release Manager

Data Manager

Test Manager



# MLOps Excite

## Organization: Communication


Image rights are held by: Dennis Wegener, Fraunhofer IAIS

# MLOps Excite
# The MLOps Cycle

# MLOps Excite

## The MLOps Cycle: Exploration

### Exploration is explanation-oriented

- Almost all components of the CRISP-DM/ Notebook as unstructured processing of DS projects
- A uniform development environment for all Data Scientists (Dev-Container) is advisable
- Central data layer to become independent of local storage in the environments of the Data Scientists

### Notebook Versioning & Testing

- Central versioning is a must-have for any software development
- Testing of notebooks already possible here!
- Optional: MLFlow as a tracking tool also in the manual area of data analysis
- MLFlow for comparing model runs and storing models

**Tools:**

Jupyter

Docker

mlflow

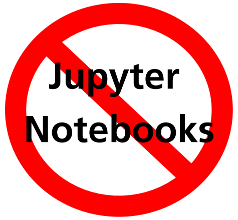
Fraunhofer IAIS

# MLOps Excite

## The MLOps Cycle: Development

### Development is structure- and efficiency-oriented

- Transition to structured source code
- Comparison of notebooks vs. source code
- Modularization of code (software engineering standards)




- Relevant here for the first time:
  - **The model is not the result!**
    There is a "usage shell" (e.g. web service, pipeline) around it.

- Preparation of release planning (git versioning) at the latest here.


Image rights are held by: Dennis Wegener, Fraunhofer IAIS

# MLOps Excite

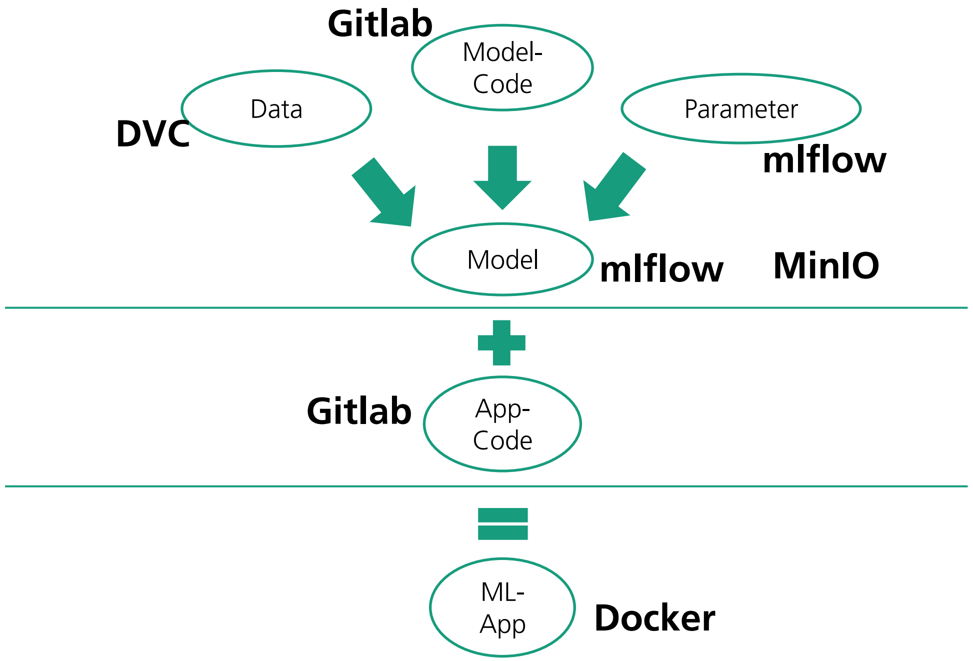
## The MLOps Cycle: Development

### Versioning and deployment

- Versioning of source code "business as usual“
- Now additionally necessary: Versioning of (ML) artefacts
- Multi-layer development
  - Code is now linked to models, data and parameters
  - Traceability of training runs
  - In addition, the model is part of an application

- Will become much more essential in the next step to Continuous Integration
  - ML pipelines become relevant

- Provision of artefacts in containerized environment (independent of dev environment)



# MLOps Excite

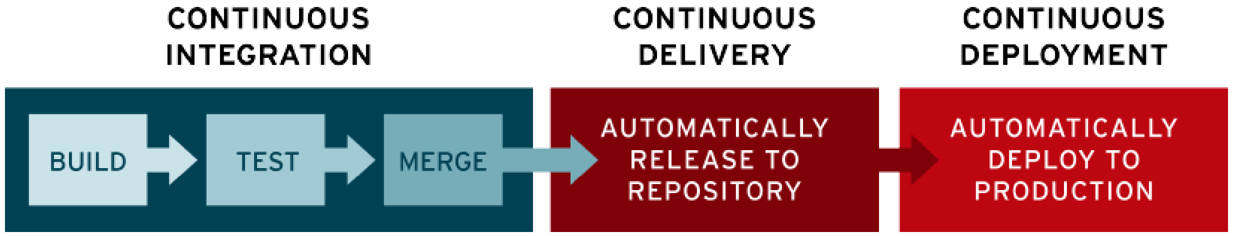
## The MLOps Cycle: CI/CD

### Continuous Integration and Continuous Deployment

- CI – Continuous Integration
  - Automation in developer work (code changes, merging, building, testing)
- CD – Continuous Delivery / Deployment
  - Automation during delivery (push into repository, rollout after staging/production)
- Different phases of the software pipeline
- Different levels of automation
- Tooling: Atlassian Bamboo, Gitlab, Jenkins, Airflow ....

**Tools:**

Pytest
Docker
Gitlab
mlflow
Kubernetes
Harbor
OpenShift




https://www.redhat.com/de/topics/devops/what-cicd-pipeline

# MLOps Excite

## The MLOps Cycle: Operation and Monitoring

### Recognizing change and deriving decisions

- Important principle:
  - "Models learn from the past, are supposed to represent reality and are supposed to make predictions about the future"
  - But: Reality changes!
- It is therefore important to observe models in use and to detect changes (drift detection)
  - User feedback (structured recording or ad-hoc)
  - Technical monitoring of the application (error logs, availability, ...)
  - Monitoring of the model itself (model metrics, performance measures)
- All these metrics need to be considered already during development
- Define measurement scales and set limits
- Decide on retraining, application adaptation (feature request, bug fixing, etc.)

**Tools:**

**Nagios**

**Ganglia**

**Grafana**

**Prometheus**

# MLOps Excite

## The MLOps Cycle: Review

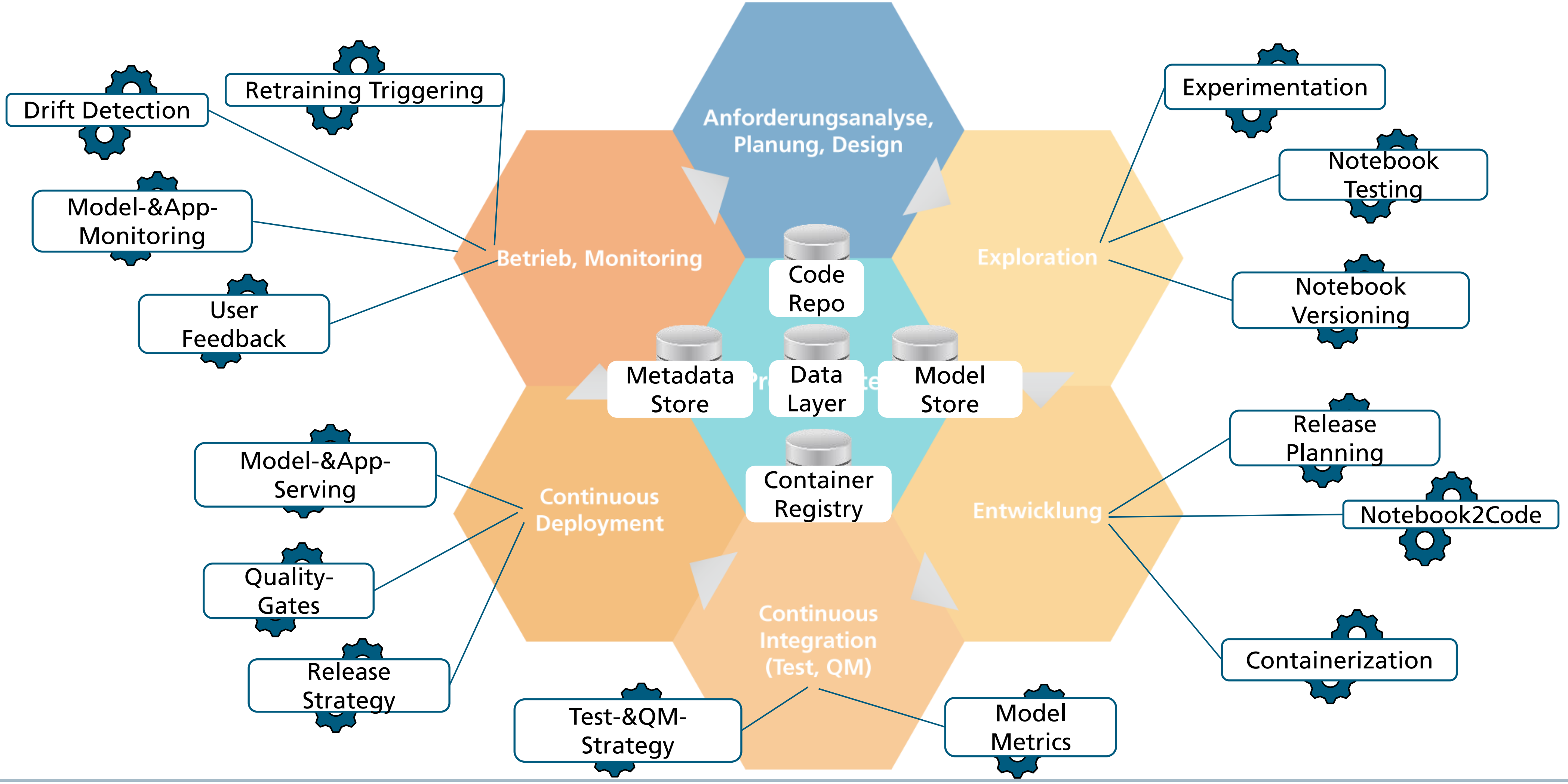



Image rights are held by: Dennis Wegener, Fraunhofer IAIS

# MLOps Excite
# ML Pipelines and Automation

# MLOps Excite

## ML Pipelines and Automation

### What exactly is meant by automation?

- We consider automation as
  - Running series of experiments during exploration
  - Transferring pipelines from exploration to production
  - Creation of production models
  - Monitoring & re-training of models

- Demarcation - we do not do AutoML
  - AutoML (Automated Machine Learning) enables non-experts to use machine learning models and techniques without data science knowledge by automation and simplification of parts of the ML pipeline

# MLOps Excite

## ML Pipelines and Automation: Automation levels

### MLOps Level 0

- Manual processing of data, training, evaluation, deployment, production

### MLOps Level 1

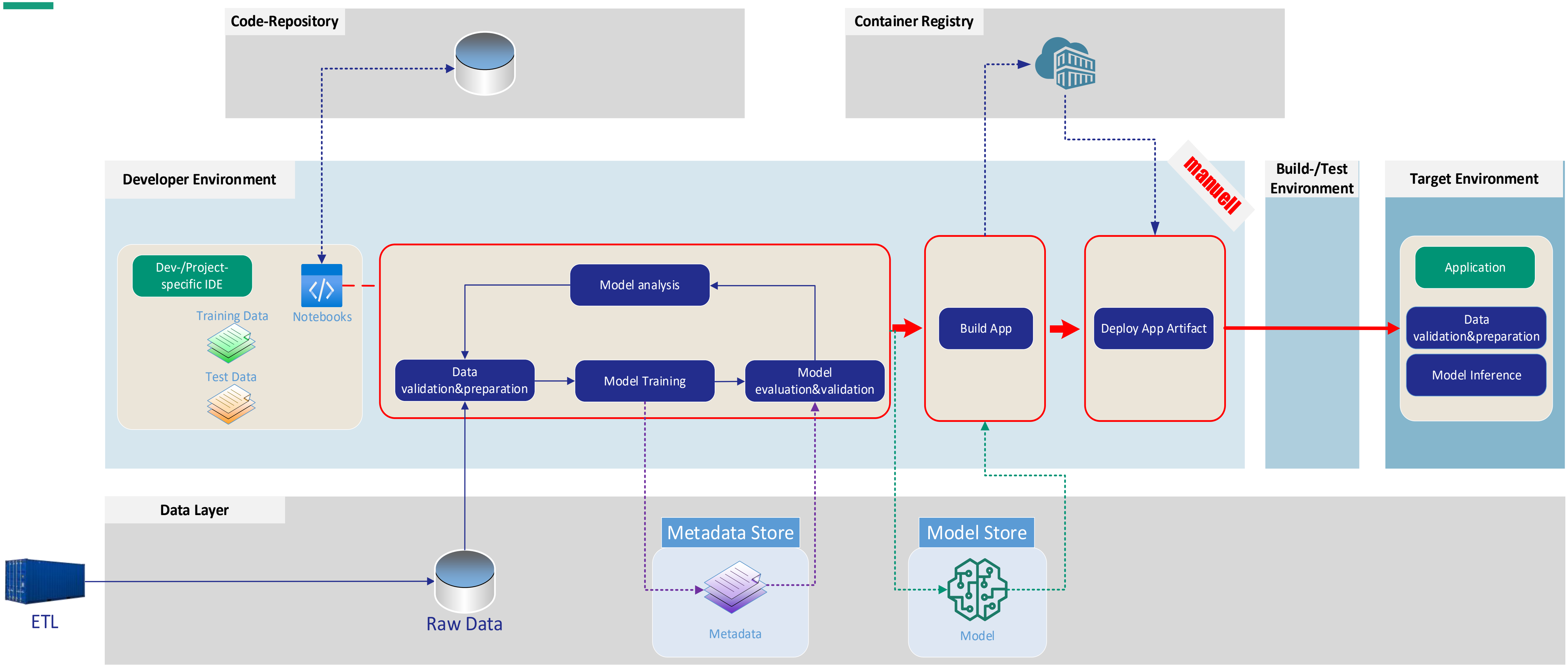
- Versioning of ML scripts, artefacts
- Containerization of all components
- Automation of training and evaluation
- Central data storage
- Definition of product-internal responsibilities and processes

### MLOps Level 2

- Automated deployment (in acceptance environment)
- ML in production with logging and metrics
- Transfer of product-internal roles and processes into organization

# MLOps Excite

## ML Pipelines and Automation: Automation levels








Fraunhofer IAIS

# MLOps Excite

## ML Pipelines and Automation: Automation levels

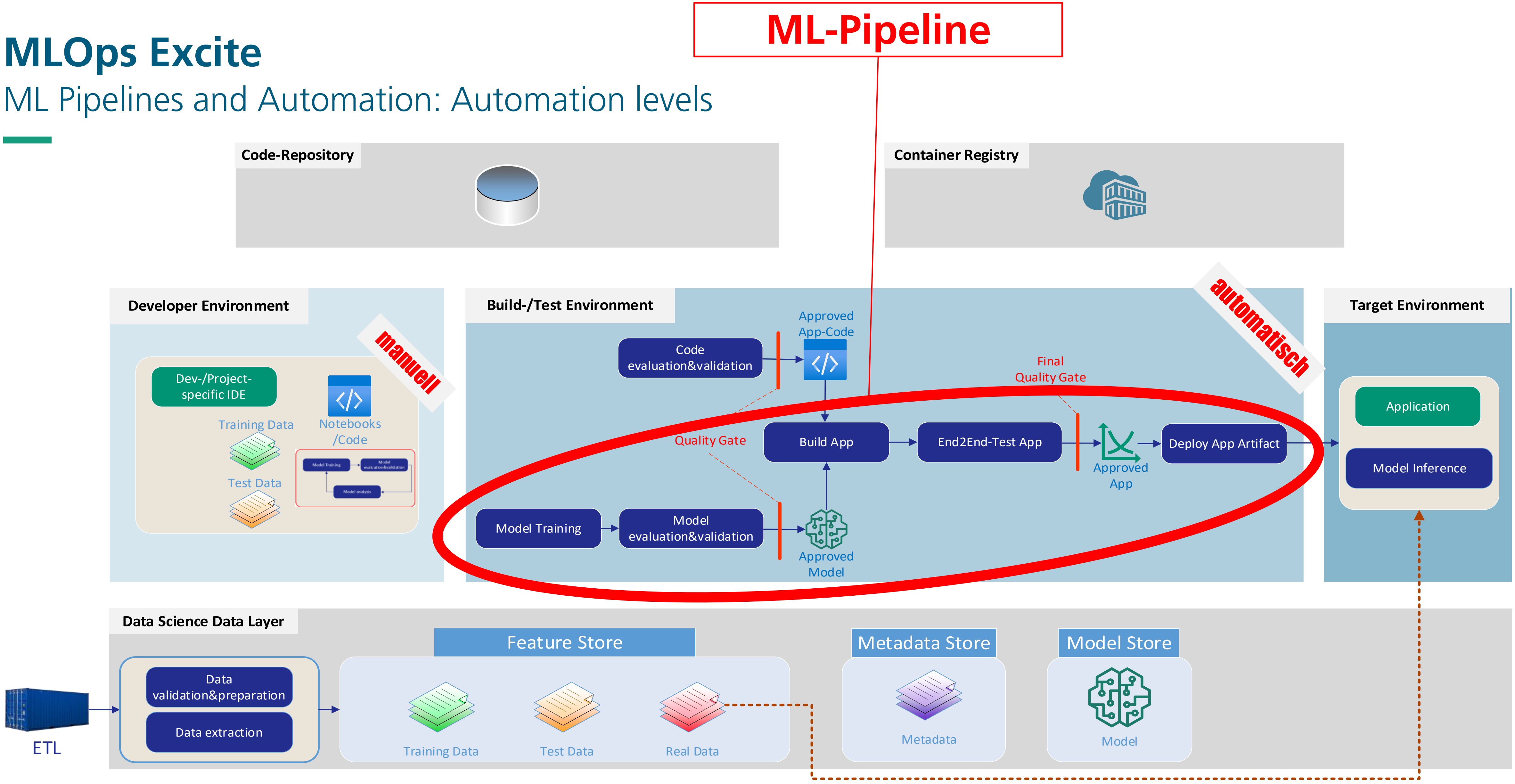



Image rights are held by: Dennis Wegener, Fraunhofer IAIS

Fraunhofer IAIS

# MLOps Excite

## ML Pipelines and Automation: Automation levels

**Monitoring and Continuous Training**

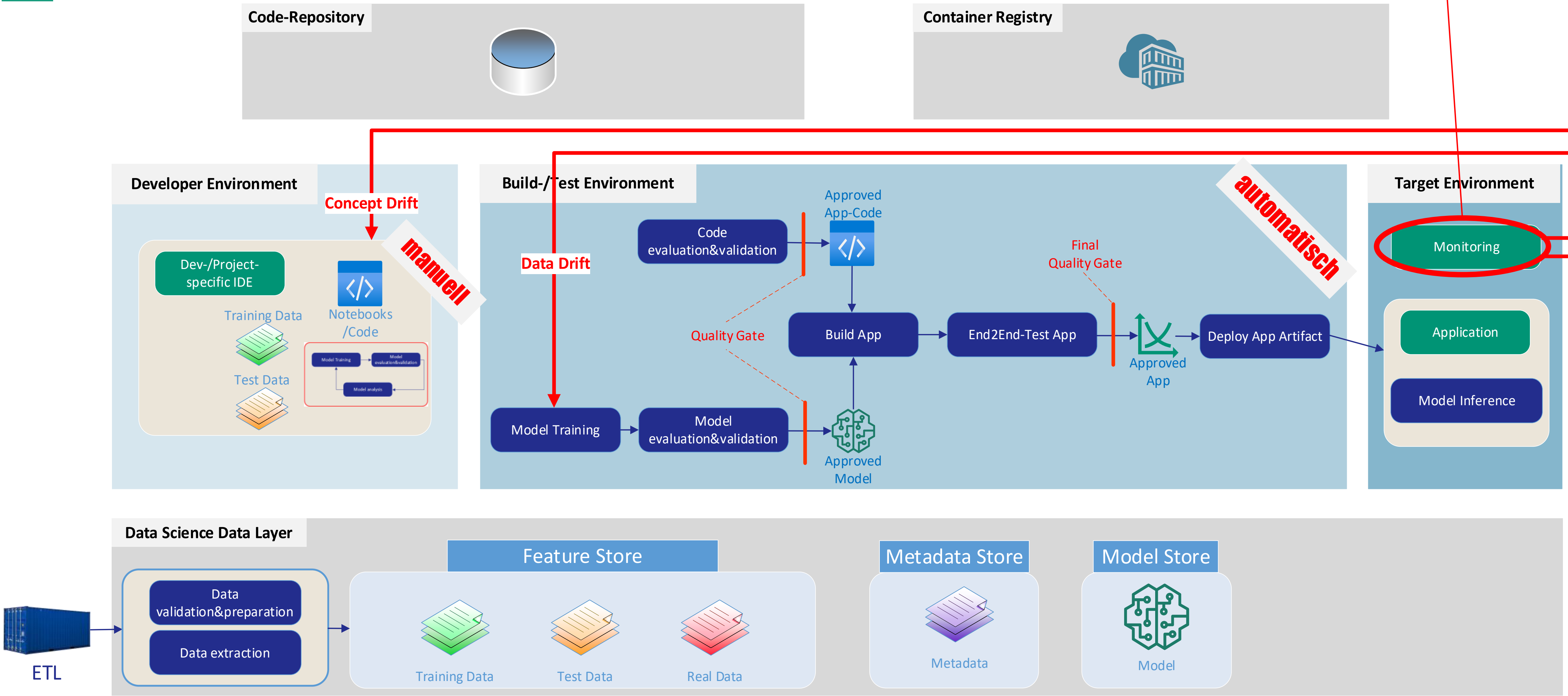



Image rights are held by: Dennis Wegener, Fraunhofer IAIS



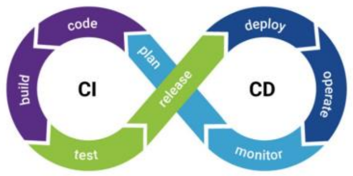
Fraunhofer IAIS

# MLOps Excite

## ML Pipelines and Automation

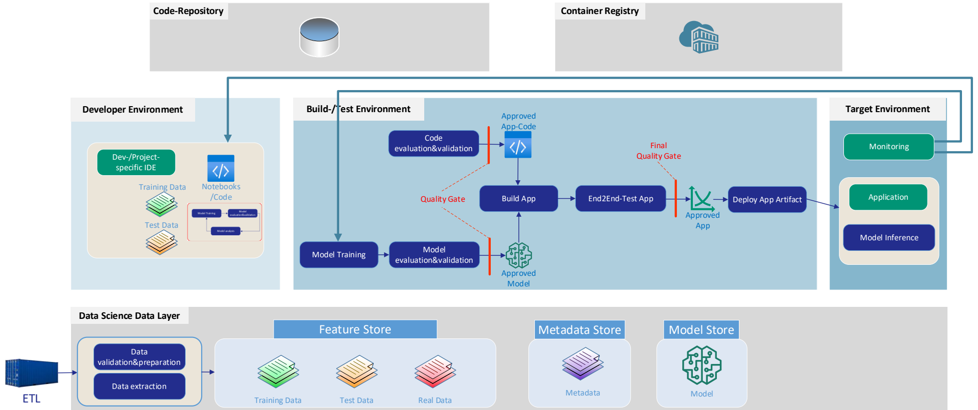
### DevOps + ML = MLOps

- **ML pipelines** need data
  - Data pipeline (Clean, Verify, Tag, ...)
  - Different data pots (train, test, validation)
- **ML Pipelines** include ML specific steps
  - Training Pipeline
  - Model artefact (architecture, parameters, weights,...)
- **ML Pipelines** or artefacts are deployed in ML-specific environments
  - TensorFlow, PyTorch, ...
- **ML Pipelines** cover large parts of the ML Cycle and need to cache results
  - Tracking and evaluation of experiments
  - Reproducibility & traceability (data input, predictions, performance metrics)
  - Quality assurance (more than unit testing, up to AB testing)
  - Monitoring in production




https://journeyofquality.com/2020/11/20/23-ci-cd-tools/








Fraunhofer IAIS

# MLOps Excite
# Become a student assistant at IAIS-MLOps

# MLOps Excite

## Become a student assistant at IAIS-MLOps

### Various possible opportunities for students

- **Participation in industry projects**
  - Taking responsibility
- **Developing state-of-the-art technology in research projects**
  - Contribute to publications
- **Bachelor and master theses**
  - Scientific projects
  - Industry projects

# Student Assistant Job

## Become a student assistant at IAIS-NLU

### Initiativbewerbung / Speculative application

- **QR-Code or Link**
  - https://jobs.fraunhofer.de/job/Sankt-Augustin-Initiativbewerbung-studentische-Hilfskraft-53757/765891401/
- Apply for one of our Teams „NLU“ or „MLOps“
- Or visit this link to get an overview
  - https://www.iais.fraunhofer.de/de/karriere.html
  - https://www.iais.fraunhofer.de/en/career.html

# Machine Learning Operations
# MLOps Excite

**Team MLOps**