Choosing the right MLOps Platform: Key Capabilities & Model Monitoring Insights

L.Berberi, V. Kozlov, B. E Sanchis, K. Alibabaei, L. Duda, G. Molto, G. Nguyen, J. Sainz-Pardo Diaz, V. Tran, A. Calatrava, A. Garcia

lisana.berberi@kit.edu
Karlsruhe Institute of Technology

Outline

Outline

Funded by the European Union

Outline

Funded by the European Union

Outline

1 MLOps definition and components

2 MLOps platforms/tools

3 Model Performance Monitoring

4 Key takeaways

Funded by the European Union

| MLOps definition and components |

Funded by
the European Union

DevOps vs MLOps

DevOps

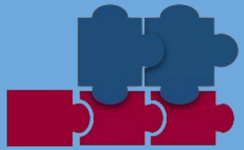
set of **practices and tools** to streamline the **software development lifecycle**

code | infra | env

MLOps

set of **practices and tools** to streamline the **machine learning model lifecycle**

data | model



code | infra | env

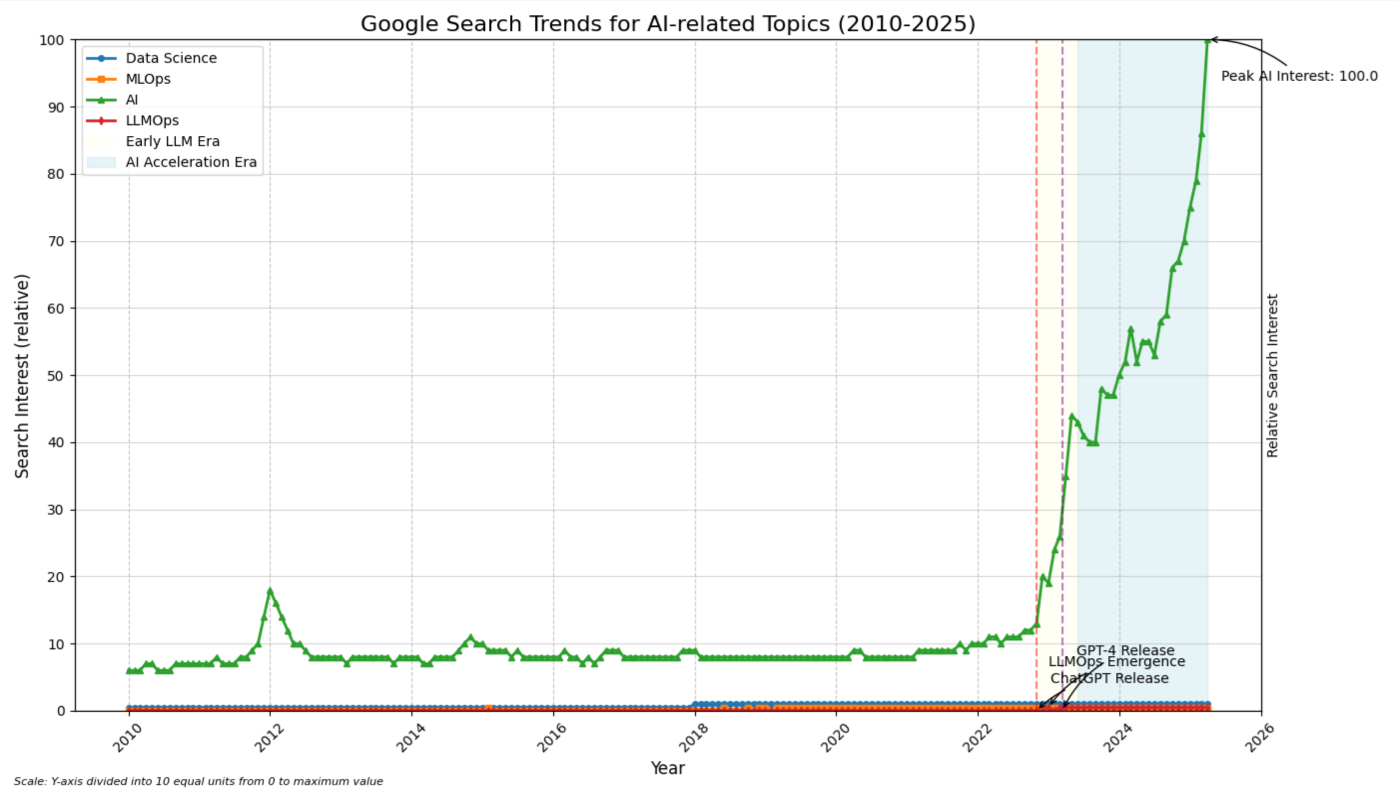
Funded by the European Union

DevOps vs MLOps

Common Goals

Automation
CI/CD
Reliability
Scalability

History

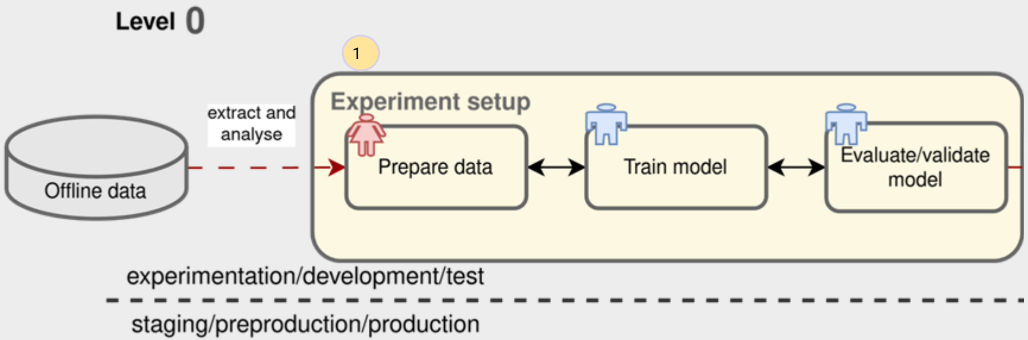


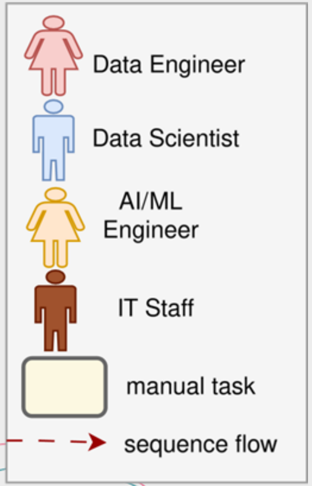
Funded by the European Union

MLOps components

Adapted from Google
([MLOps: Continuous delivery and automation pipelines in machine learning)](MLOps: Continuous delivery and automation pipelines in machine learning)

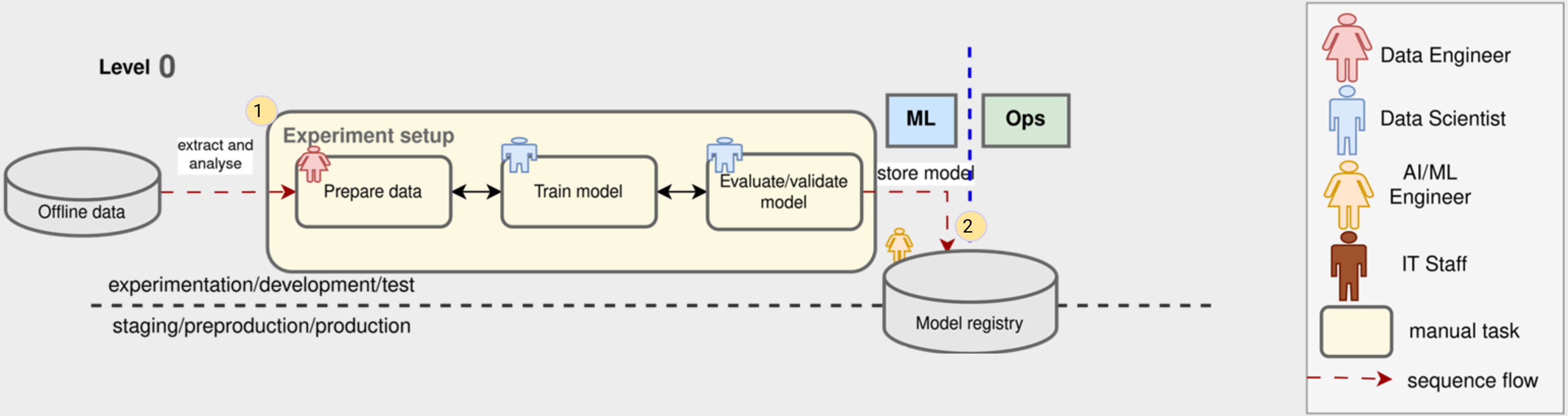
Scan me!





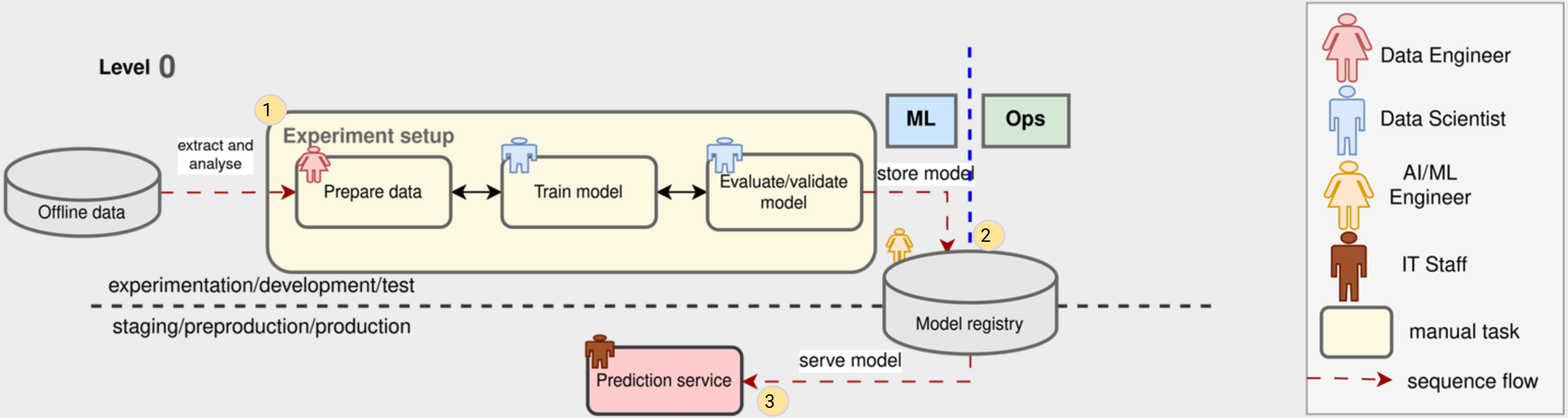
Funded by the European Union

MLOps components

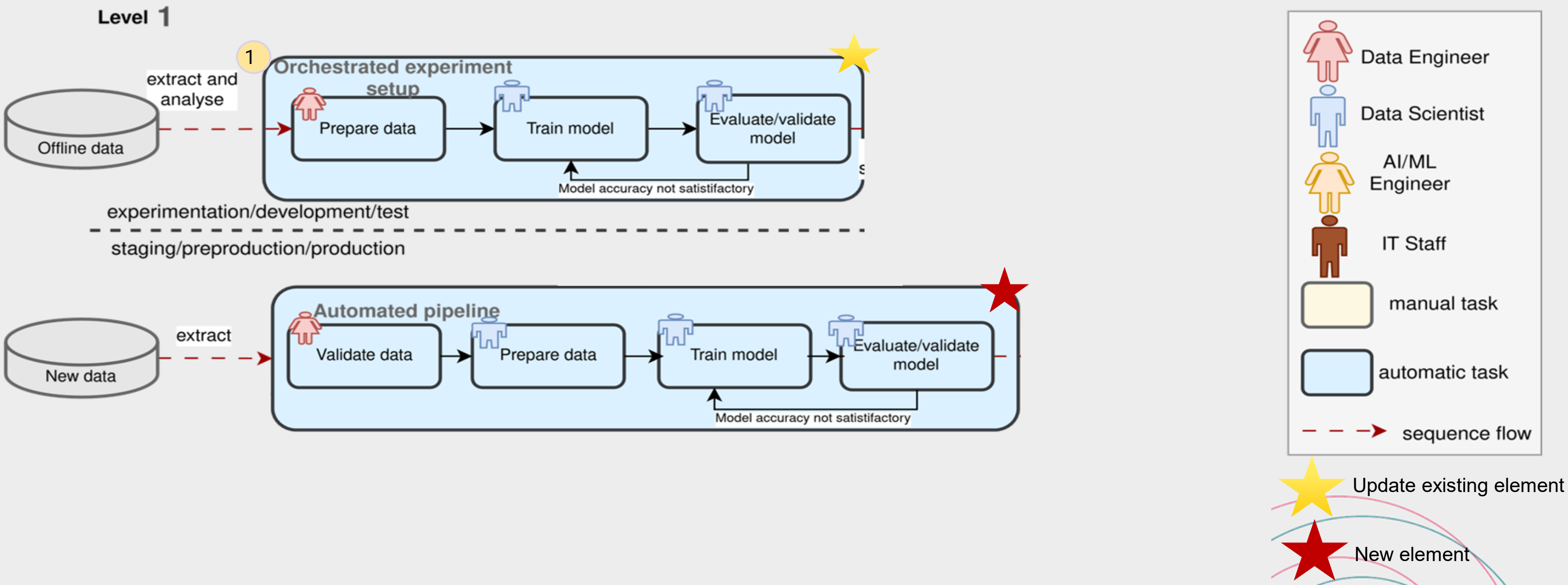


Funded by the European Union

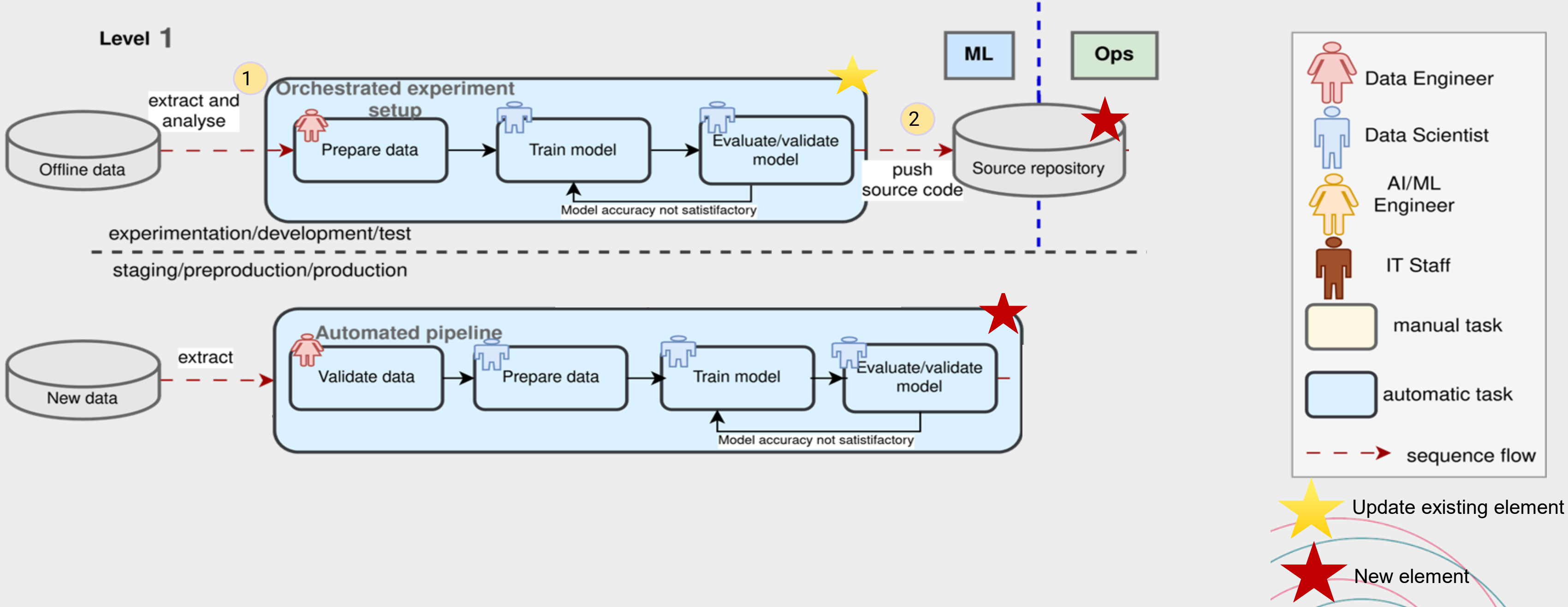
MLOps components



MLOps components

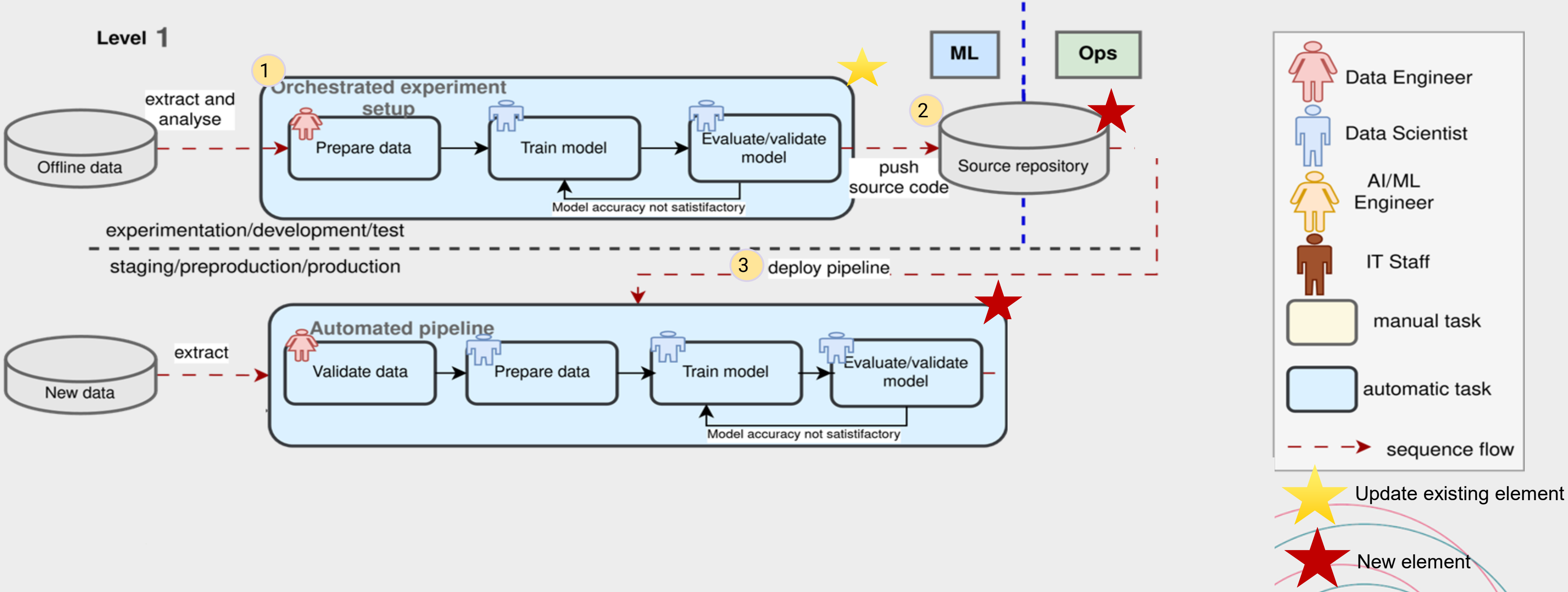


MLOps components



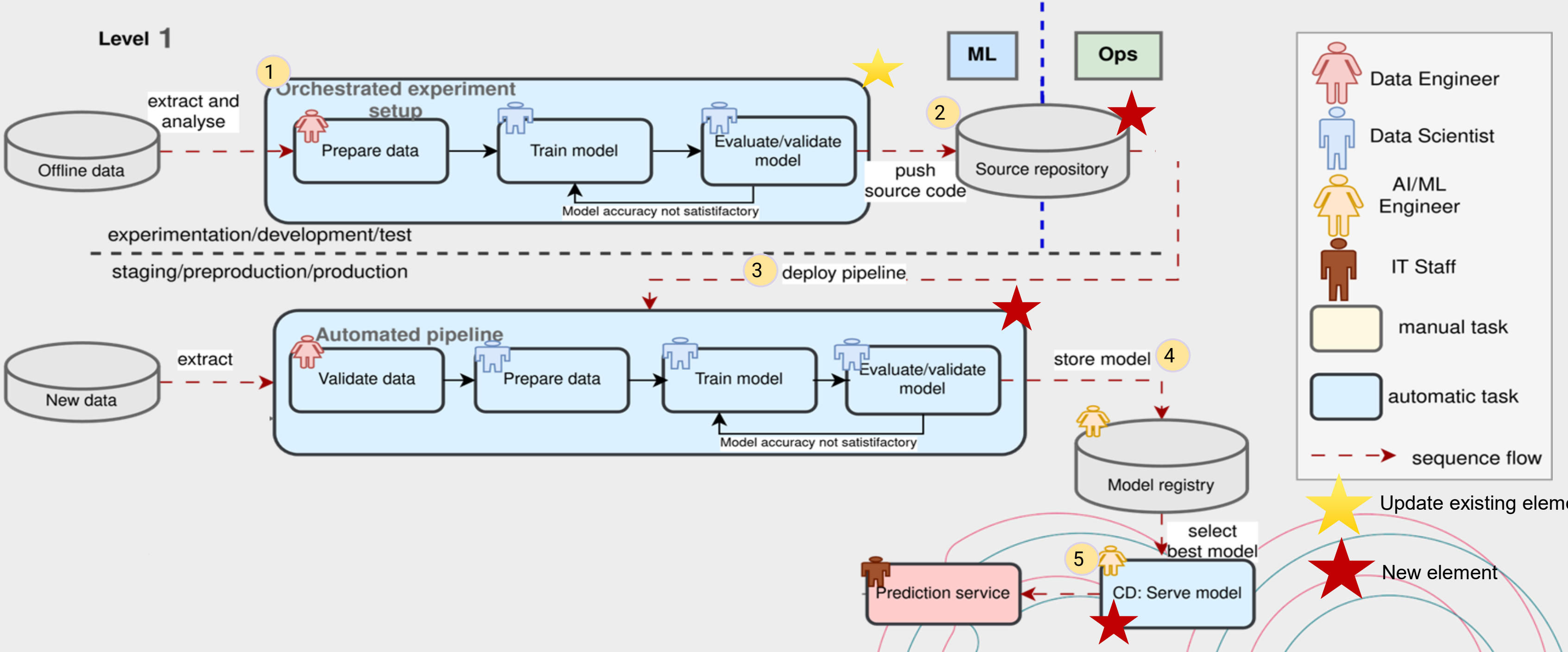
Funded by the European Union

MLOps components



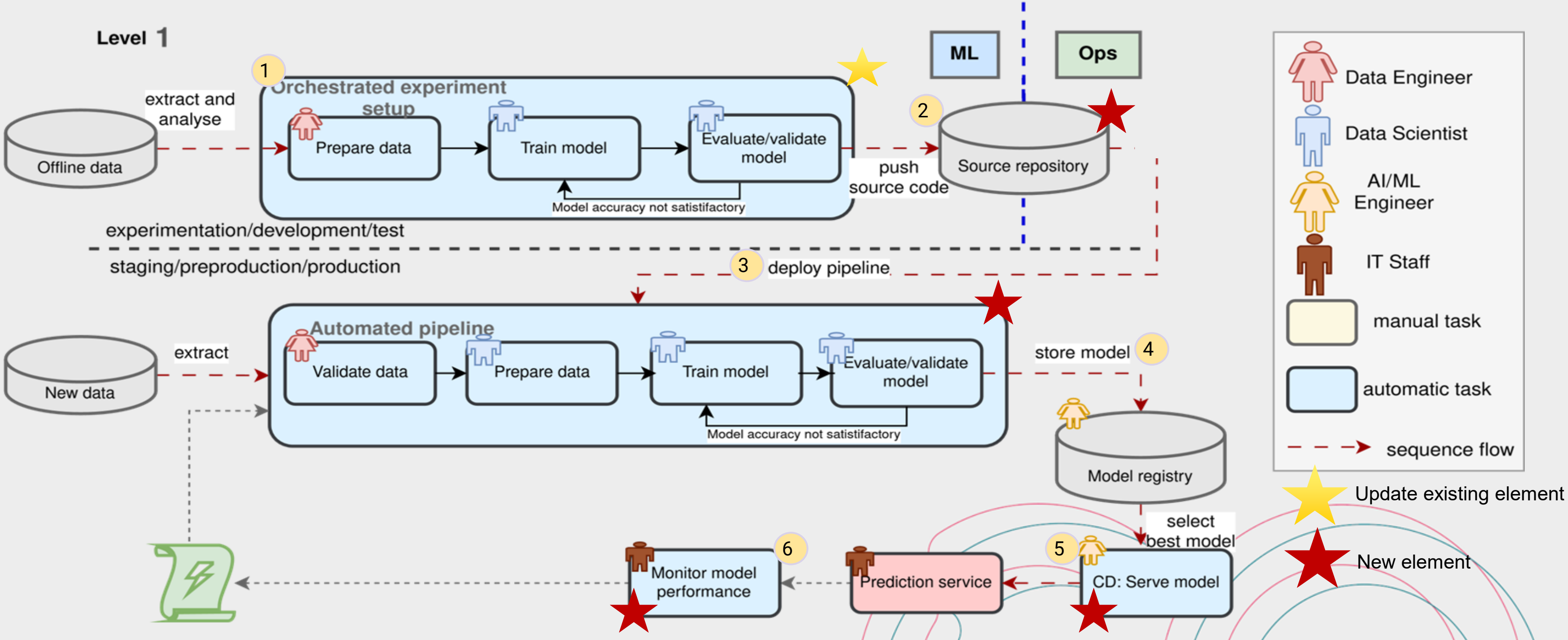
Funded by the European Union

MLOps components

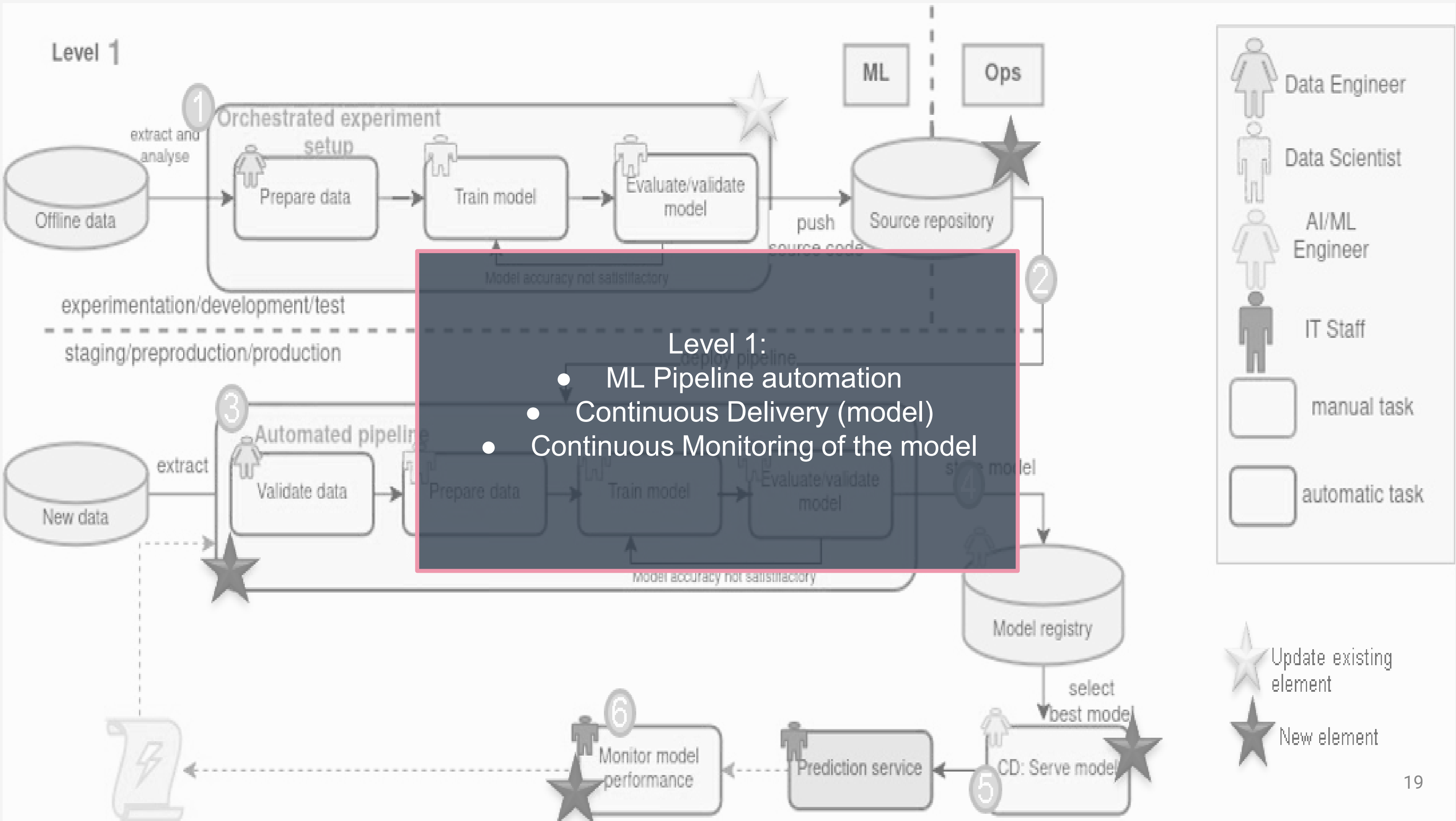


Funded by
the European Union

MLOps components



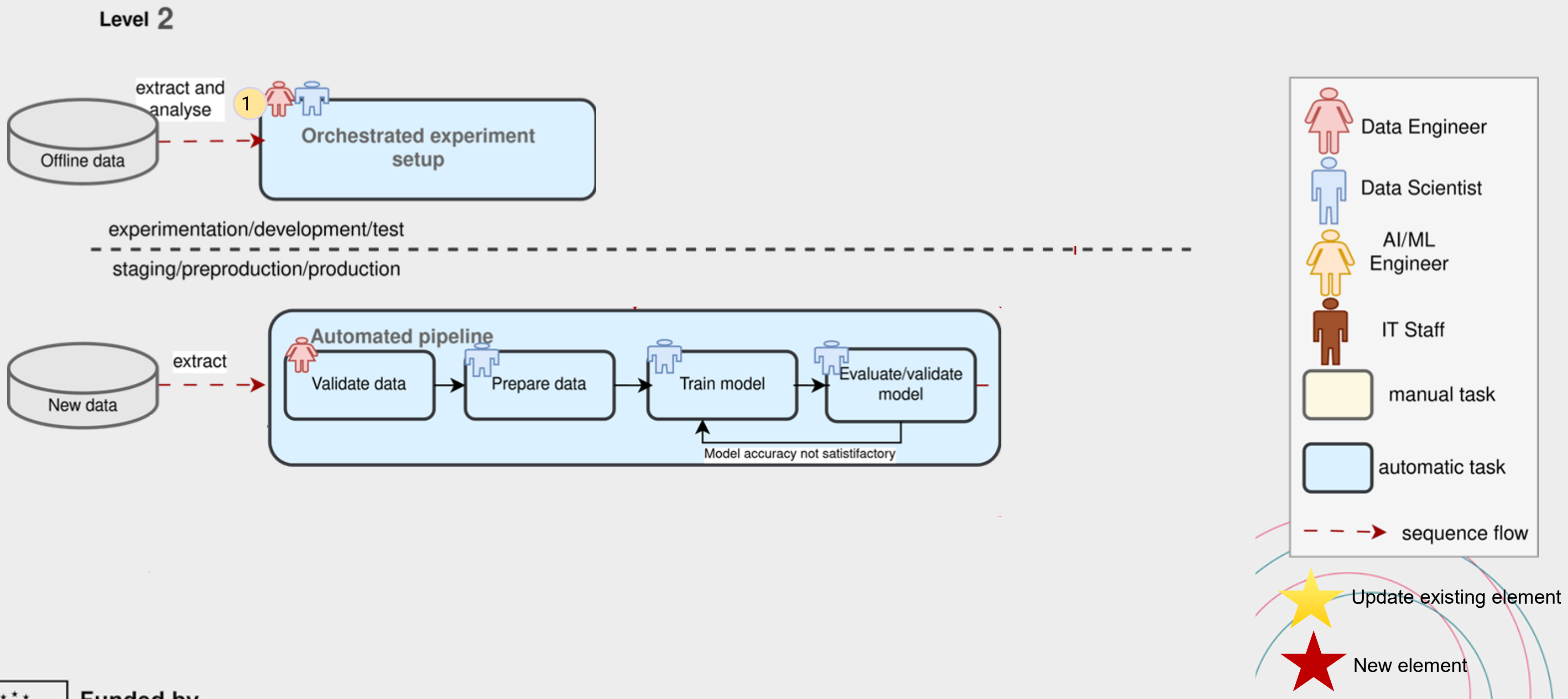
Funded by the European Union



Level 1:

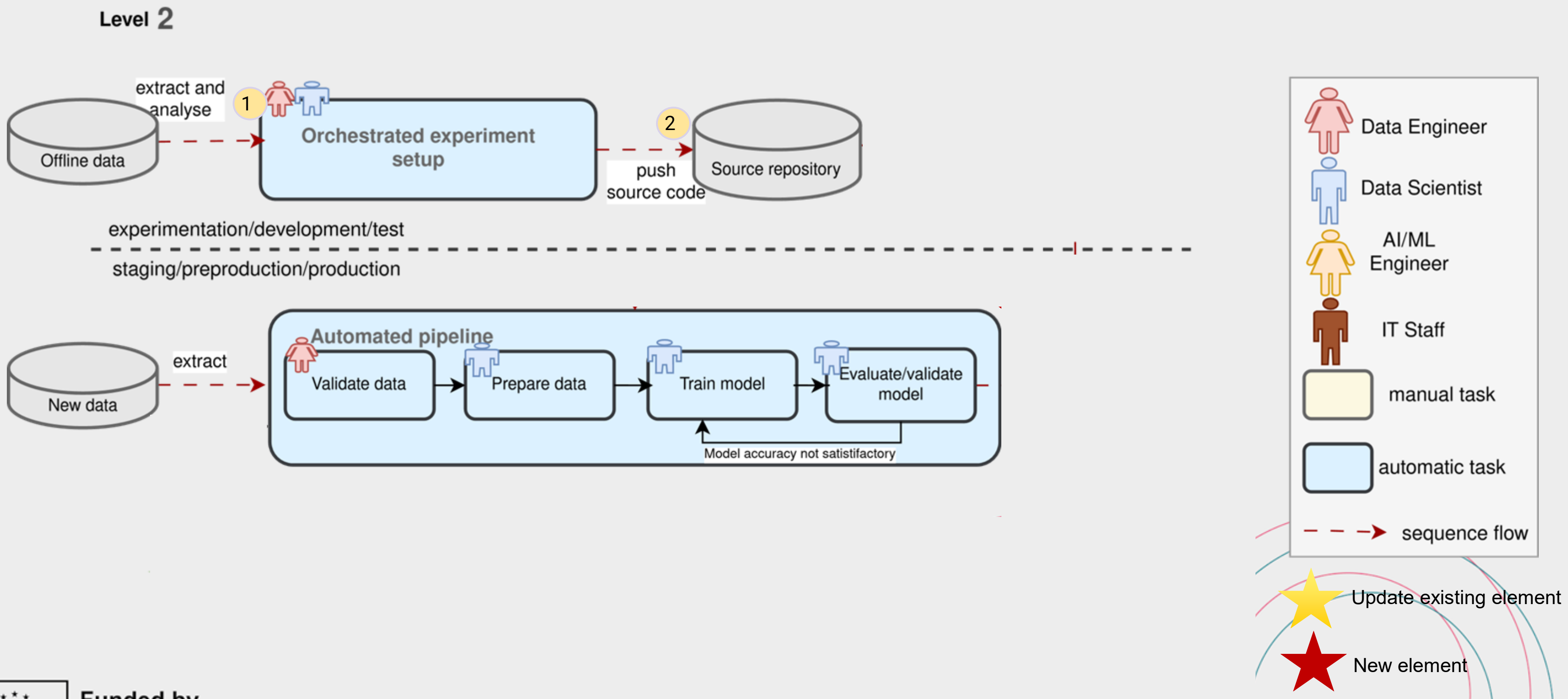
- ML Pipeline automation
- Continuous Delivery (model)
- Continuous Monitoring of the model

MLOps components



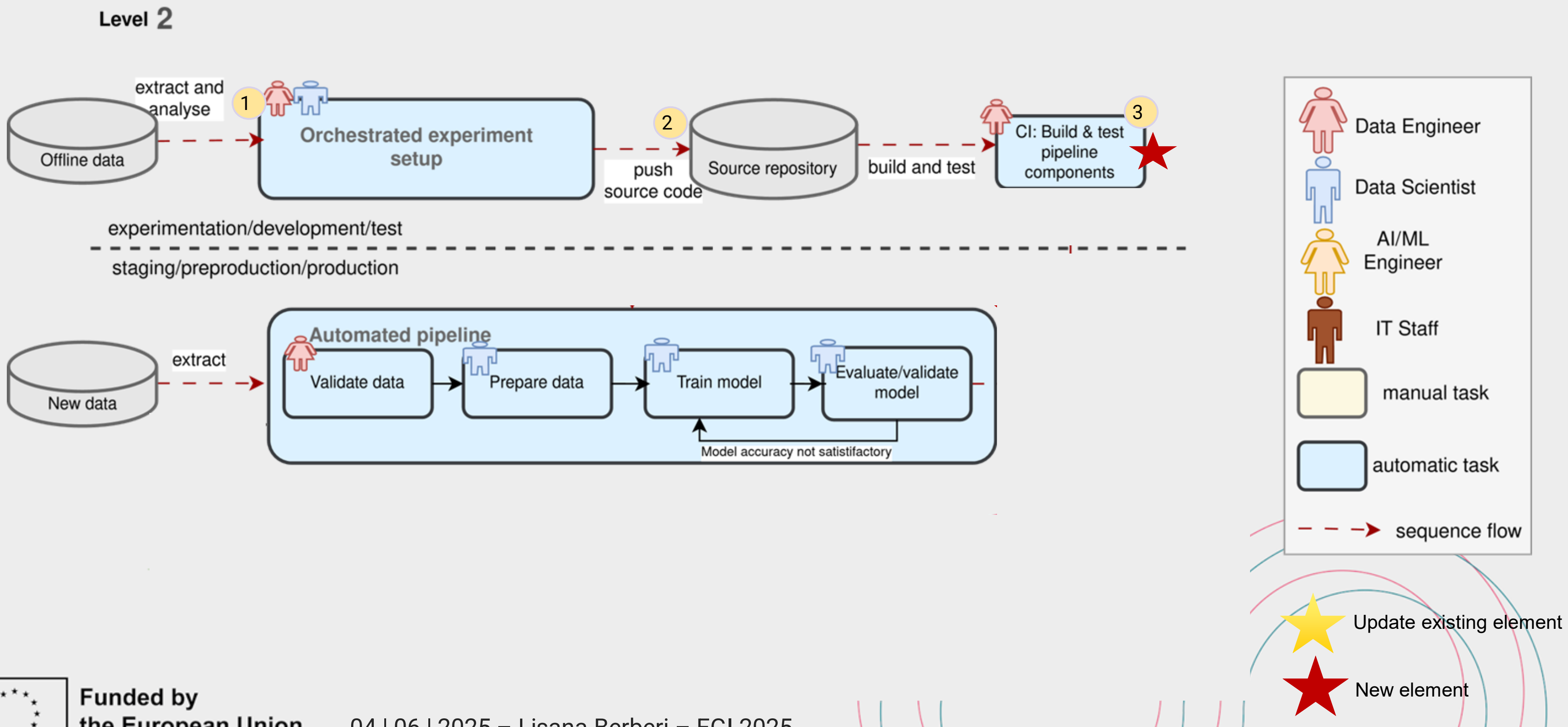
Funded by the European Union

MLOps components

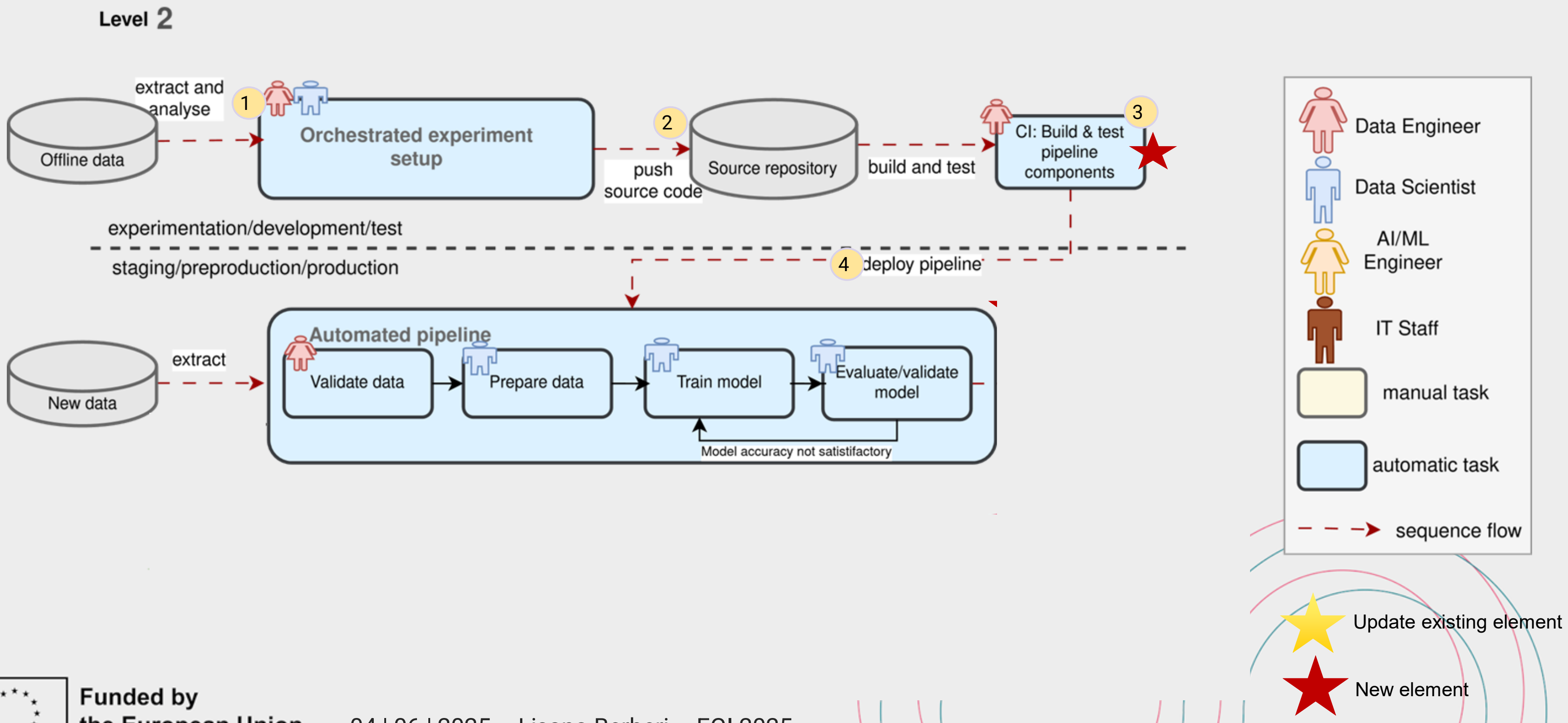


Funded by the European Union

MLOps components

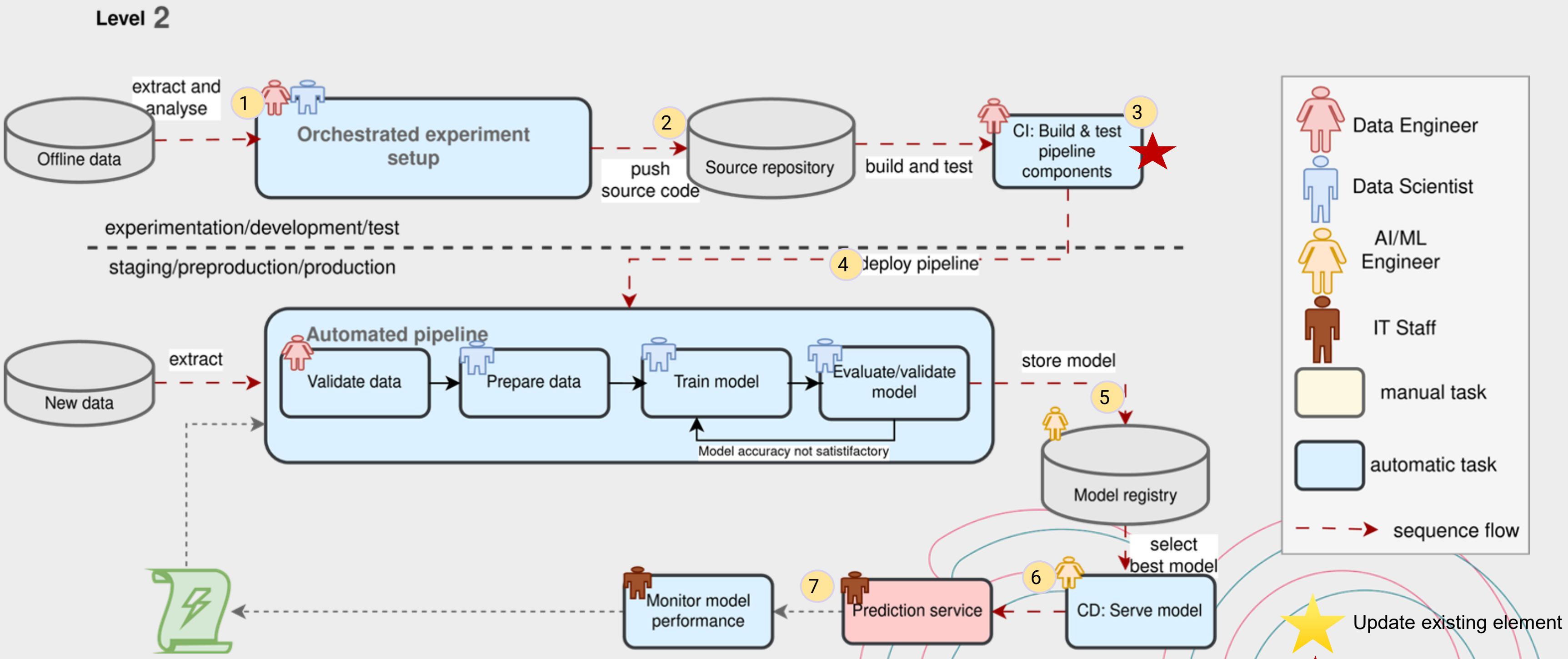


MLOps components



Funded by the European Union

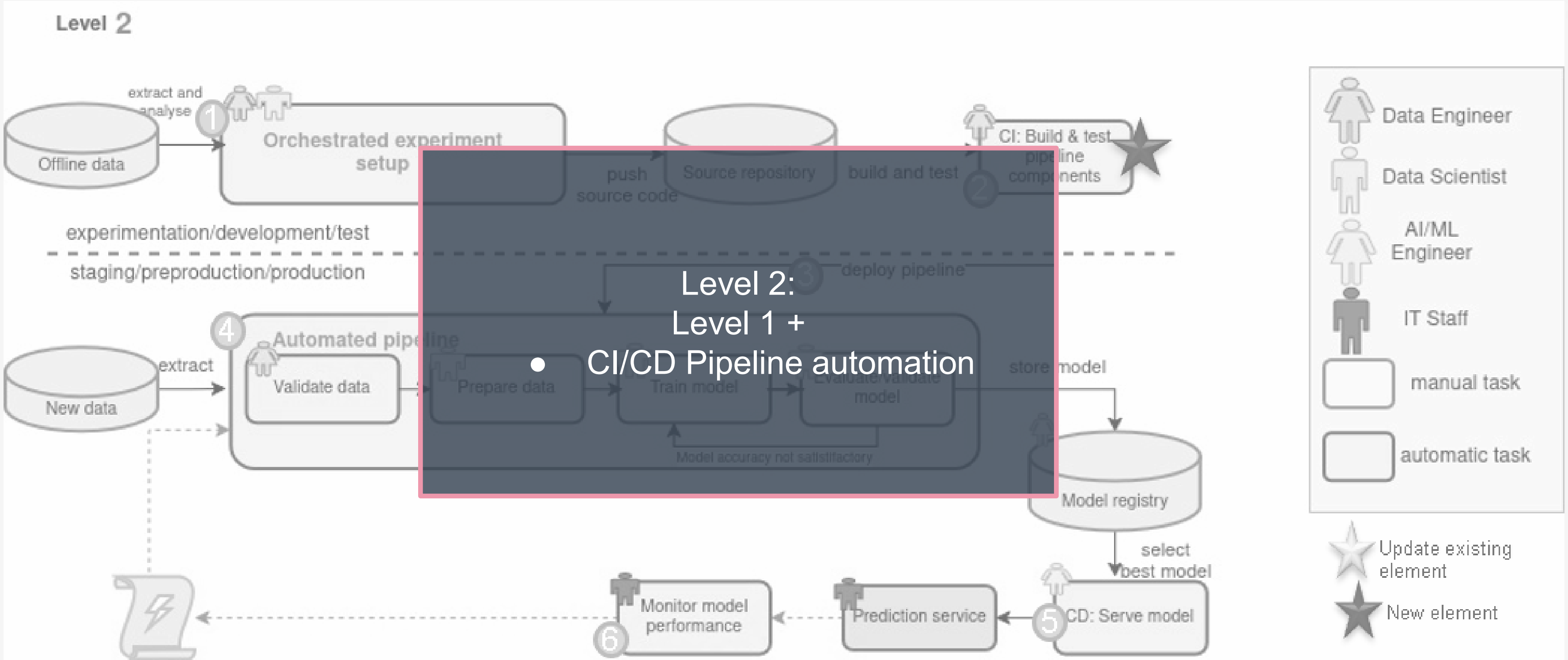
MLOps components



Funded by the European Union

AI4 eosc

MLOps components



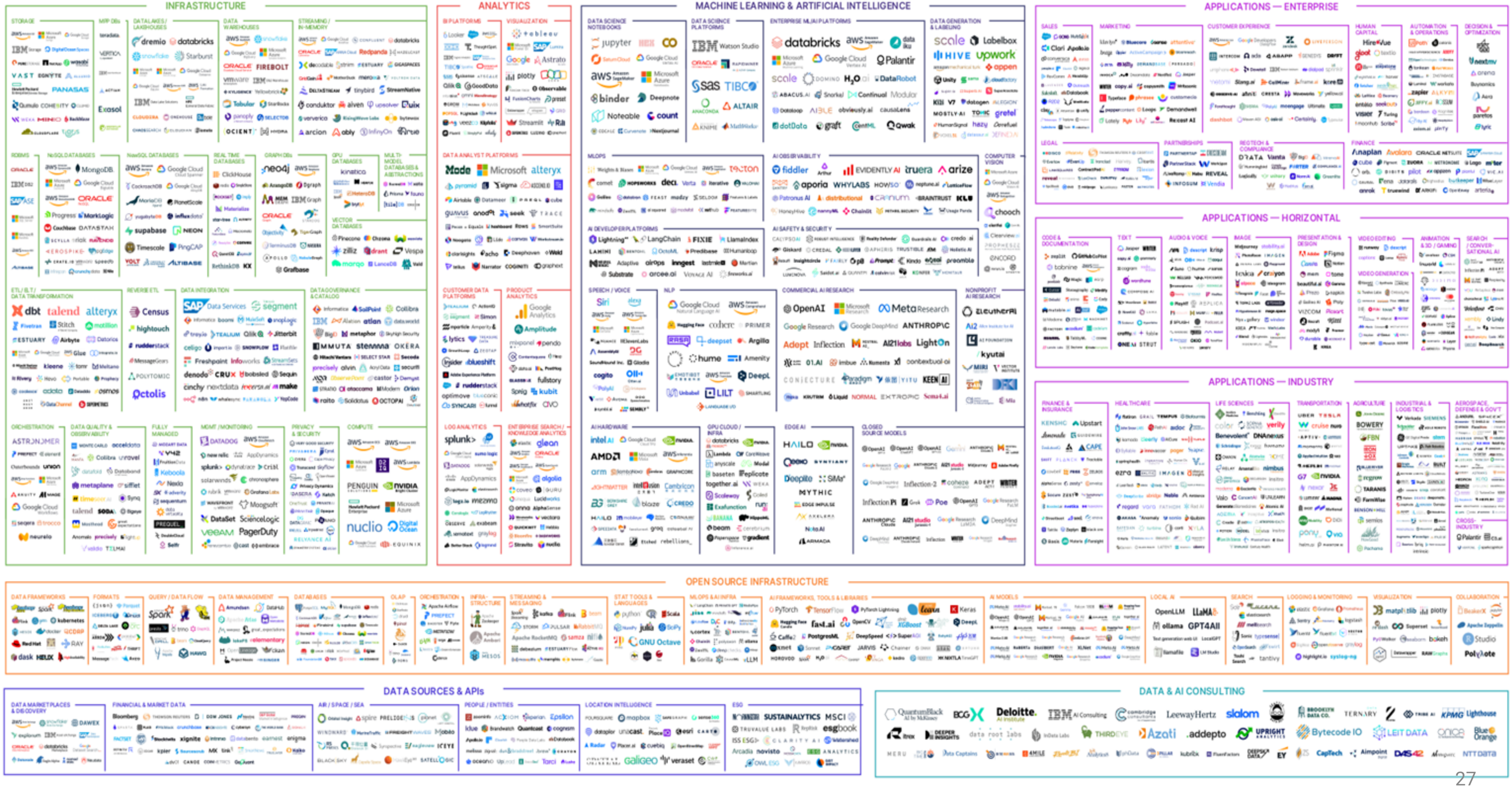
Level 2:
Level 1 +
- CI/CD Pipeline automation

Funded by the European Union

| MLOps Platform/Tools |

Funded by
the European Union

THE 2024 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE



Version 1.0 - March 2024 © Matt Turck (@mattturck), Aman Kabeer (@AmanKabeer11) & FirstMark (@firstmarkcap) Blog post: mattturck.com/MAD2024 Interactive version: MAD.firstmarkcap.com Comments? Email MAD2024@firstmarkcap.com

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Scan me!

Step 1

Feature/ Capability Analysis

- Evaluated 16 MLOps open source platforms across core capabilities.
- 10 capabilities drawn from the AI-Infrastructure Report (2023) and academic literature.
- Focus: Experiment tracking, model development, orchestration etc.

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Step 1

Feature/ Capability Analysis

- Evaluated 16 MLOps open source platforms across core capabilities.
- 10 capabilities drawn from the AI-Infrastructure Report (2023) and academic literature.
- Focus: Experiment tracking, model development, orchestration etc.

Table 3 Notable open-source MLOps platforms

Product	**GitHub Stars**	**O** Orchestration	**DT** Distributed Training	**CM** Code Management	**MDV** Model Development	**MTV** Model Testing/Validation	**MI** Model Inference	**MDP** Model Deployment	**ETMS** Experiment Tracking and Metadata Store	**DVM** Data Versioning and Management	**MPM** Model Performance Monitoring	**Full** Score	**Partial** Score
MLflow	19 K			✓			✓✓	✓✓	✓✓			30%	10%
Prefect	17.7 K	✓✓		✓						✓✓		20%	10%
Kubeflow	14.5 K	✓✓	✓✓	✓	✓✓	✓	✓✓	✓✓	✓✓			60%	20%
Dagster	12 K	✓✓					✓	✓✓		✓✓		30%	10%
W&B (WB)	9.2 K	✓	✓✓	✓✓	✓✓	✓✓		✓	✓✓	✓✓	✓✓	70%	10%
MetaFlow	8.3 K	✓✓	✓✓	✓					✓✓	✓	✓	20%	30%
Mage	8 K	✓✓					✓✓	✓	✓	✓✓	✓	30%	30%
Pachyderm	6.2 K	✓✓	✓	✓✓	✓✓	✓✓	✓✓	✓	✓	✓✓		60%	30%
Flyte	5.8 K	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓		90%	0%
ClearML	5.7 K	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	100%	0%
Seldon core	4.4 K	✓✓				✓✓	✓✓	✓✓	✓		✓✓	50%	10%
ZenML	4.2 K	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	100%	0%
Polyaxon	3.6 K	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓	✓✓	90%	10%
TFX	2.1 K	✓✓	✓✓	✓✓	✓✓	✓	✓✓	✓✓	✓	✓✓		70%	20%
MLeap	1.5 K	✓✓					✓✓	✓✓			✓	30%	10%
MLRun	1.5 K	✓✓	✓✓	✓	✓✓	✓	✓✓	✓✓	✓✓	✓✓	✓✓	80%	20%

Funded by the European Union

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

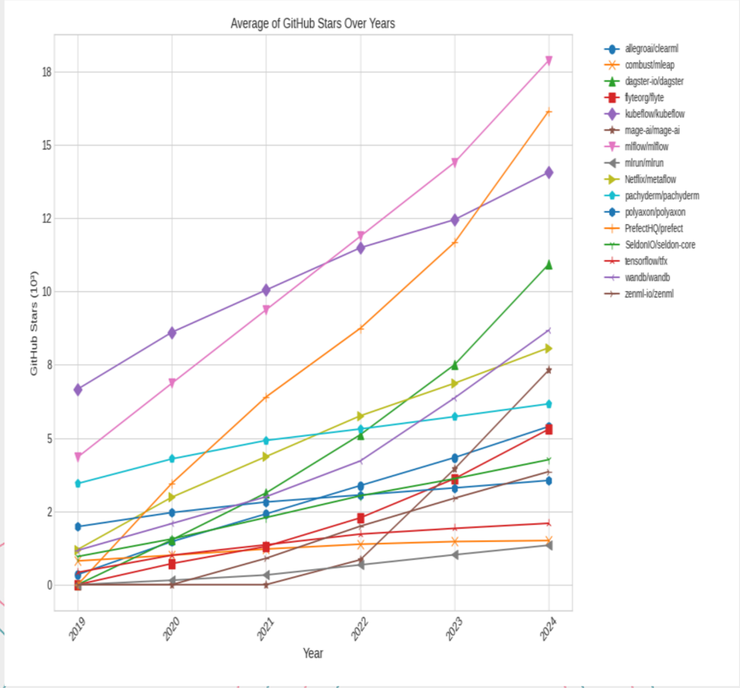
Step 1	Step 2
Feature/ Capability Analysis	GitHub Stars Growth Assessment
- Evaluated 16 MLOps open source platforms across core capabilities. - 10 capabilities drawn from the AI-Infrastructure Report (2023) and academic literature. - Focus: Experiment tracking, model development, orchestration etc.	- Assessed average GitHub stars growth over the last years. - Goal: Highlight developer interest and adoption trends

Funded by the European Union

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Step 1	Step 2
Feature/ Capability Analysis	GitHub Stars Growth Assessment
- Evaluated 16 MLOps open source platforms across core capabilities. - 10 capabilities drawn from the AI-Infrastructure Report (2023) and academic literature. - Focus: Experiment tracking, model development, orchestration etc.	- Assessed average GitHub stars growth over the last years. - Goal: Highlight developer interest and adoption trends



Funded by the European Union

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Step 1	Step 2	Step 3
Feature/ Capability Analysis	GitHub Stars Growth Assessment	Weighted Scoring & Feature Extraction
- Evaluated 16 MLOps open source platforms across core capabilities. - 10 capabilities drawn from the AI-Infrastructure Report (2023) and academic literature. - Focus: Experiment tracking, model development, orchestration etc.	- Assessed average GitHub stars growth over the last years. - Goal: Highlight developer interest and adoption trends	- Calculated weighted scores combining platform features and growth metrics. - List top 5 tools

Funded by the European Union

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Step 1	Step 2	Step 3
Feature/ Capability Analysis	GitHub Stars Growth Assessment	Weighted Scoring & Feature Extraction

- E
- 1 th

-Foc

Table 4 The weighted score for each product

Product	**Weight** (w_i)	**Feature-Score**	**Weighted-Score**
Kubeflow	8.89	7	**62.21**
WandB (W&B)	4.79	8	**38.30**
MLflow	10.00	3.5	**35.00**
Pachyderm	4.34	7.5	**32.55**
ClearML	2.94	10	**29.38**
Flyte	2.98	9	26.82
Polyaxon	2.65	9.5	25.22
ZenML	2.20	10	22.01
Dagster	6.13	3.5	21.44
Prefect	8.35	2.5	20.87
Mage	3.98	4.5	17.90
Metaflow	4.11	3.5	14.40
Seldon core	2.59	5.5	14.23
TFX	1.30	8	10.40
MLRun	1.00	9	9.00
MLeap	1.32	3.5	4.61

Funded by the European

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Step 1	Step 2	Step 3
Feature/ Capability Analysis	GitHub Stars Growth Assessment	Weighted Scoring & Feature Extraction

- E
- 1 th

-Foc

Table 4 The weighted score for each product

Product	Weight (w_i)	Feature-Score	Weighted-Score
Kubeflow	8.89	7 Kubernetes bound	**62.21**
WandB (W&B)	4.79	8 Limited features in the free version	**38.30**
MLflow	10.00	3.5	**35.00**
Pachyderm	4.34	7.5	**32.55**
ClearML	2.94	10	**29.38**
Flyte	2.98	9	26.82
Polyaxon	2.65	9.5	25.22
ZenML	2.20	10	22.01
Dagster	6.13	3.5	21.44
Prefect	8.35	2.5	20.87
Mage	3.98	4.5	17.90
Metaflow	4.11	3.5	14.40
Seldon core	2.59	5.5	14.23
TFX	1.30	8	10.40
MLRun	1.00	9	9.00
MLeap	1.32	3.5	4.61

d

Funded by the European

MLOps platforms/tools

- Three-step evaluation framework [L. Berberi et al. (2025)]

Step 1	Step 2	Step 3
Feature/ Capability Analysis	GitHub Stars Growth Assessment	Weighted Scoring & Feature Extraction

- E
- 1 th

-Fo

Table 4 The weighted score for each product

Product	Weight (w_i)	Feature-Score	Weighted-Score
Kubeflow	8.89	7	**62.21**
WandB (W&B)	4.79	8	**38.30**
MLflow	10.00	3.5	**35.00**
Pachyderm	4.34	7.5	**32.55**
ClearML	2.94	10	**29.38**
Flyte	2.98	9	26.82
Polyaxon	2.65	9.5	25.22
ZenML	2.20	10	22.01
Dagster	6.13	3.5	21.44
Prefect	8.35	2.5	20.87
Mage	3.98	4.5	17.90
Metaflow	4.11	3.5	14.40
Seldon core	2.59	5.5	14.23
TFX	1.30	8	10.40
MLRun	1.00	9	9.00
MLeap	1.32	3.5	4.61

Scan me!

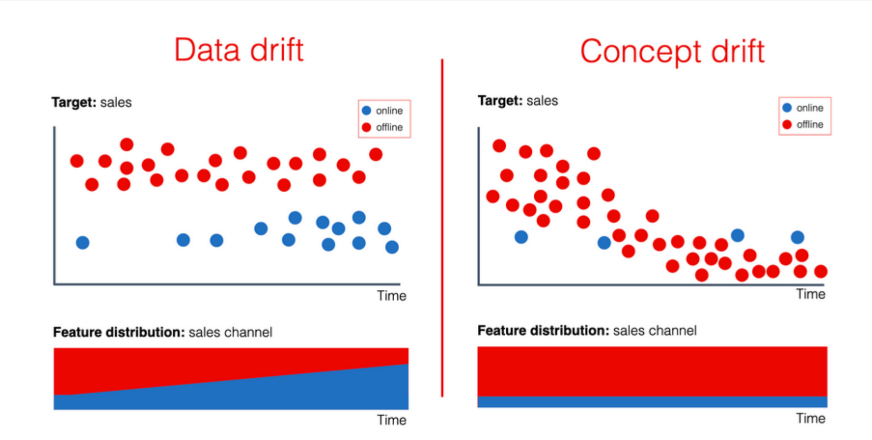
Funded by the European

| Model Performance Monitoring |

Funded by
the European Union

Drift frameworks comparison

- Data drift is a shift in the distributions of the ML model input features.
- Concept drift is a change in input-output relationships.
- **Different modes of execution:**
  - streaming: data arrives sequentially, e.g. online monitoring
  - batch: full dataset available at time of test, e.g. offline model evaluation
- Frameworks:
  - Frouros
  - River
  - Evidently
  - NannyML
  - Alibi-Detect
- Dataset:
  - Energy Data
    - Heating energy consumption data from educational buildings (schools/universities)



Data drift vs concept drift (source EvidentlyAI)

Funded by the European Union

AI4 eosc

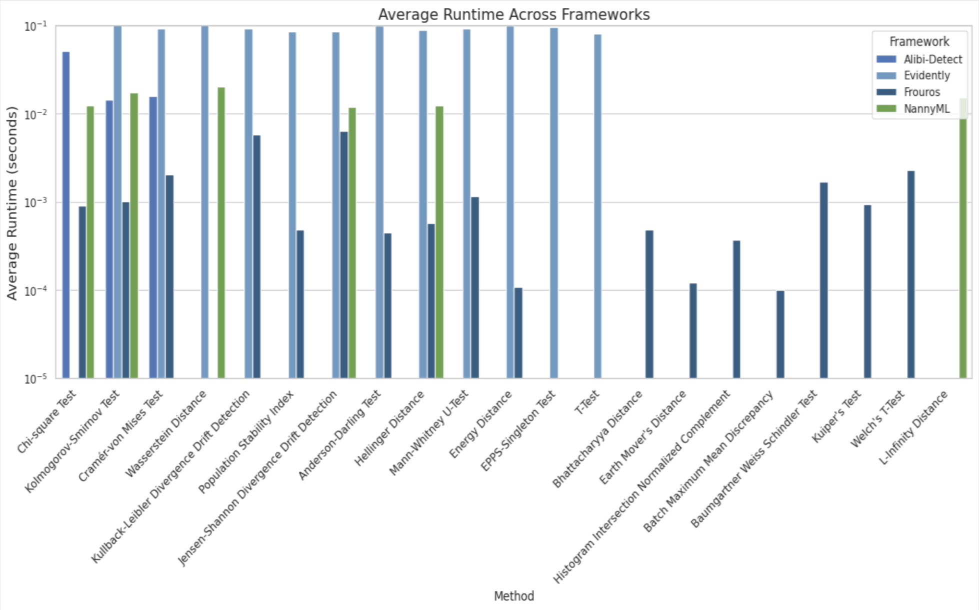
Drift frameworks comparison

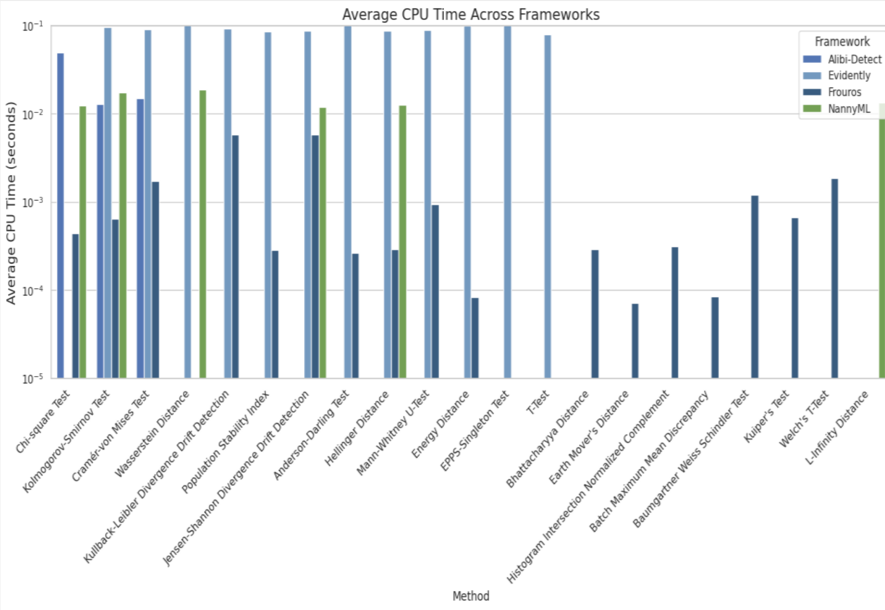
- original repo D3bench: https://github.com/mohamedyd/D3Bench
  - Publication:

Scan me!

- Extended repo: https://github.com/BorjaEst/D3Bench/tree/dev/results
- Poster:

Scan me!

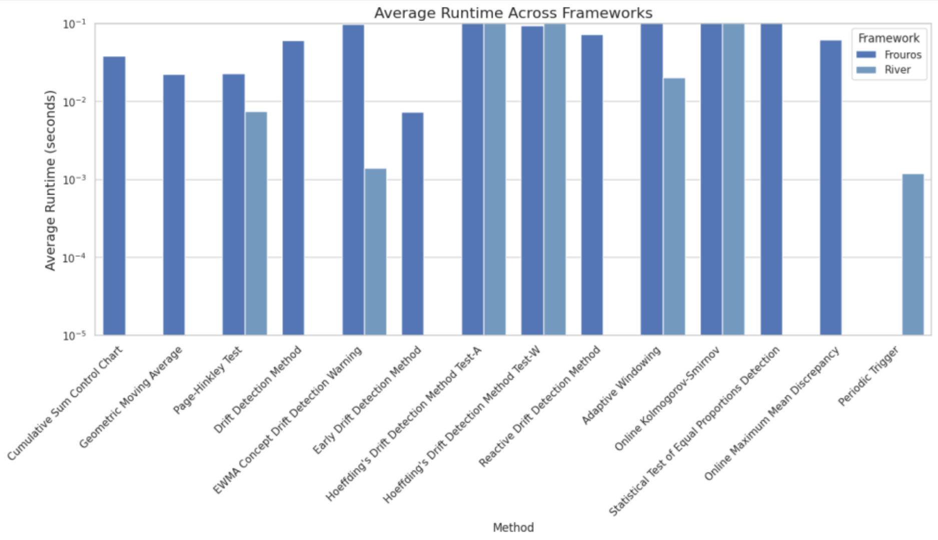


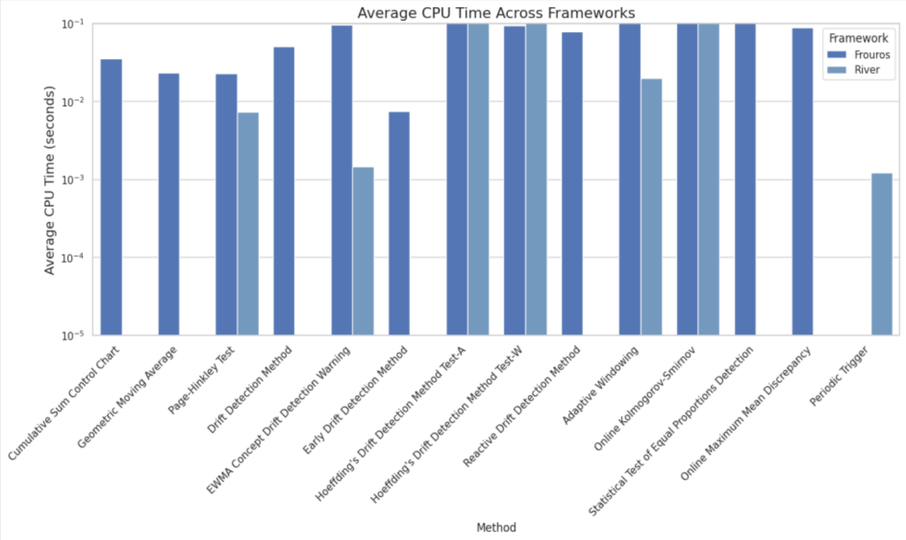


Batch Data Drift Detection methods

Drift frameworks comparison

- Tool evaluation results:





Online Concept Drift Detection methods

Funded by the European Union

| Key takeaways |

Funded by
the European Union

Key takeaways

- MLOps is critical for scaling machine learning beyond experimentation into production.
- MLOps maturity levels help assess and plan ML lifecycle automation.
- Evaluate state-of the art drift frameworks using the Extended D3Bench tool

Scan me!
MLflow cloud

Scan me!
MLflow video-1

Scan me!
MLflow video-2

Scan me!
AI4EOSC Dashboard

Scan me!
AI4EOSC MLflow Docs

Scan me!
D3bench extension

Funded by the European Union

AI4 eosc

AI4EOSC	ai4eosc-po@listas.csic.es	ai4eosc.eu

Reach us!

Thank you for your attention

The AI4EOSC consortium

Funded by
the European Union