

MLOps in Practice: Tracking and Deploying Machine Learning Models

Lisana Berberi

lisana.berberi@kit.edu

Karlsruhe Institute of Technology



Outline

1

MLOps definition and components

2

MLOps platforms/tools

3

MLflow

4

Key takeaways

Outline

- 1 MLOps definition and components
- 2 MLOps platforms/tools
- 3 MLflow
- 4 Key takeaways

Outline

1

MLOps definition and components

2

MLOps platforms/tools

3

MLflow

4

Key takeaways

Outline

1

MLOps definition and components

2

MLOps platforms/tools

3

MLflow

4

Key takeaways

| MLOps definition and components |

DevOps vs MLOps

DevOps

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

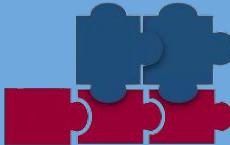


code | infra | env

MLOps

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

data | model

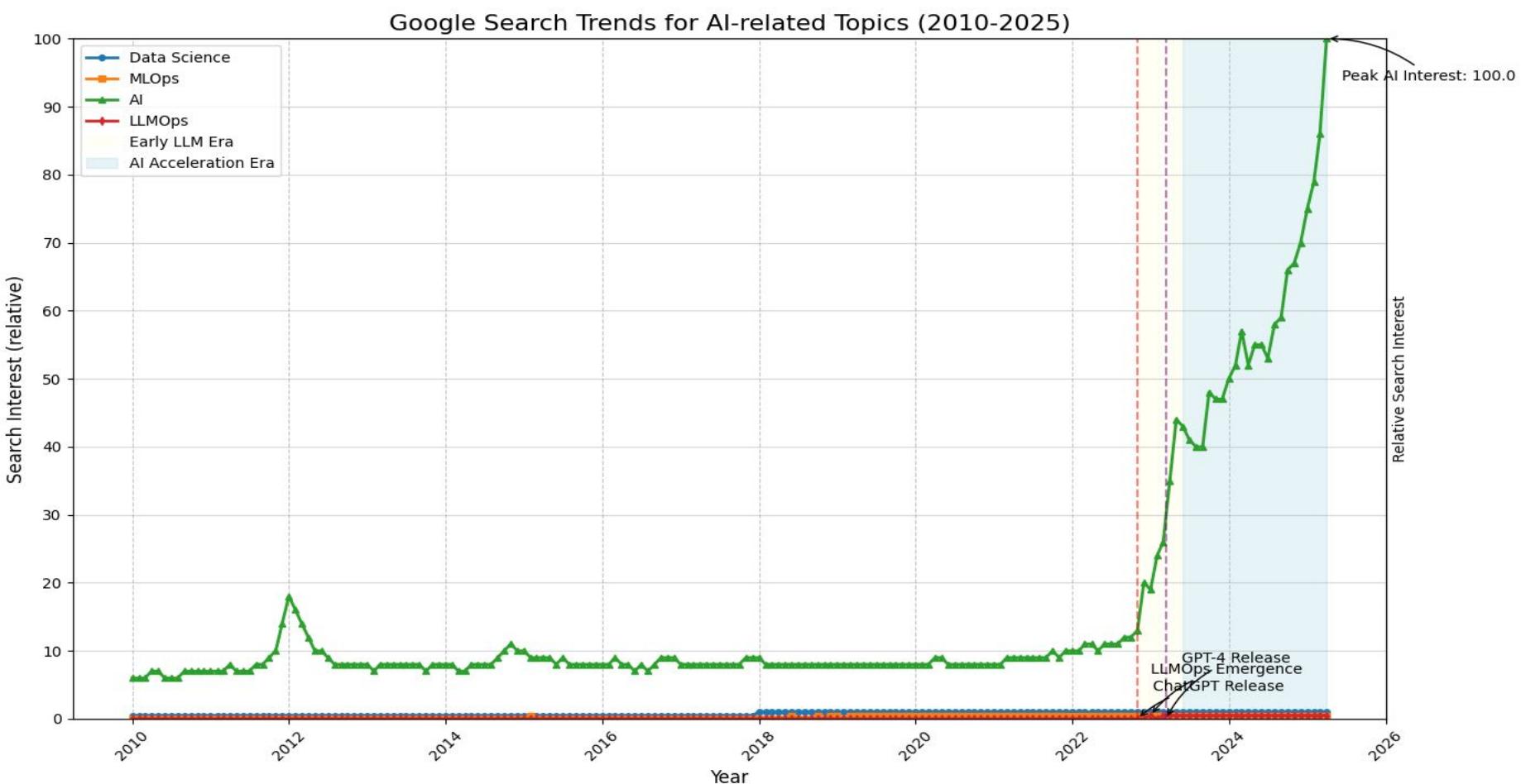


code | infra | env

DevOps vs MLOps

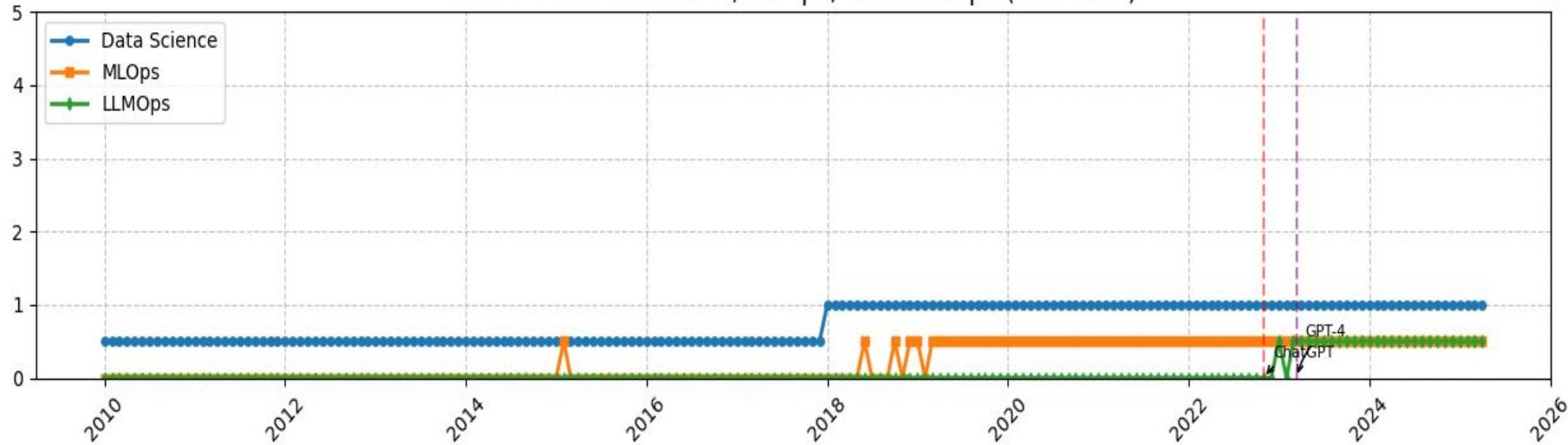


History



History

Google Search Trends: Comparison of AI, Data Science, MLOps, and LLM Ops (2010-2025)
Focus on Data Science, MLOps, and LLM Ops (0-5 scale)



| MLOps Principles |

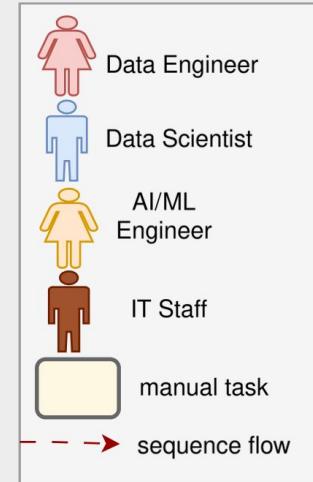
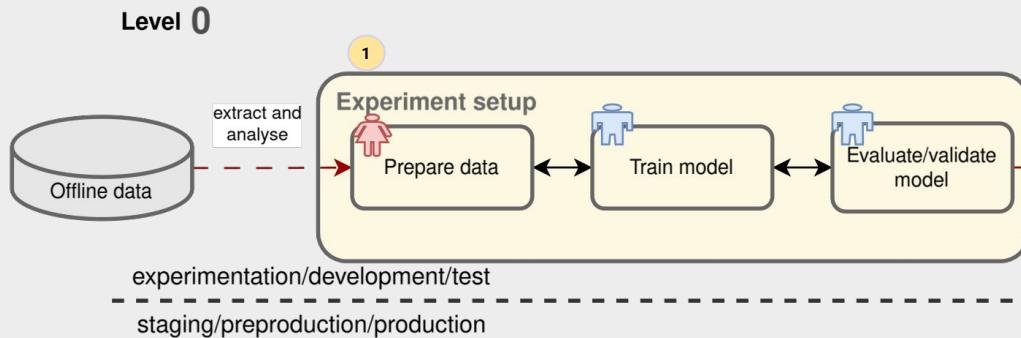
MLOps principles

- Documentation
 - for each project, you need to document business goals and KPIs;
 - ML model specifics, steps of gathering, analyzing, and cleaning data including the reasoning;
- Code quality
 - keep high-quality, maintainable code for scalable ML systems.
 - Infrastructure code quality requirements and ML model code quality requirements
- Traceability and Reproducibility
 - allows us to generate the same model with the same data and code version, guaranteeing consistency, and therefore we can roll back to the previous model version easily when needed.
 - code for gathering, analyzing, and cleaning data should be stored in a version control system.
- Monitoring
 - tracking of infrastructure costs is set up
 - for real-time inference use cases, all API requests and responses should be logged, API response time,
 - Data drift monitoring

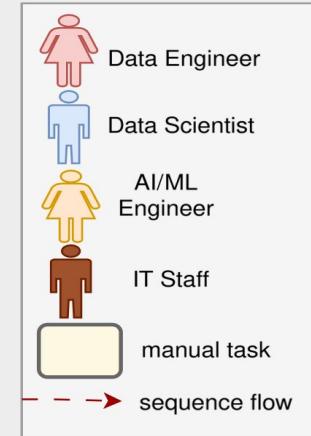
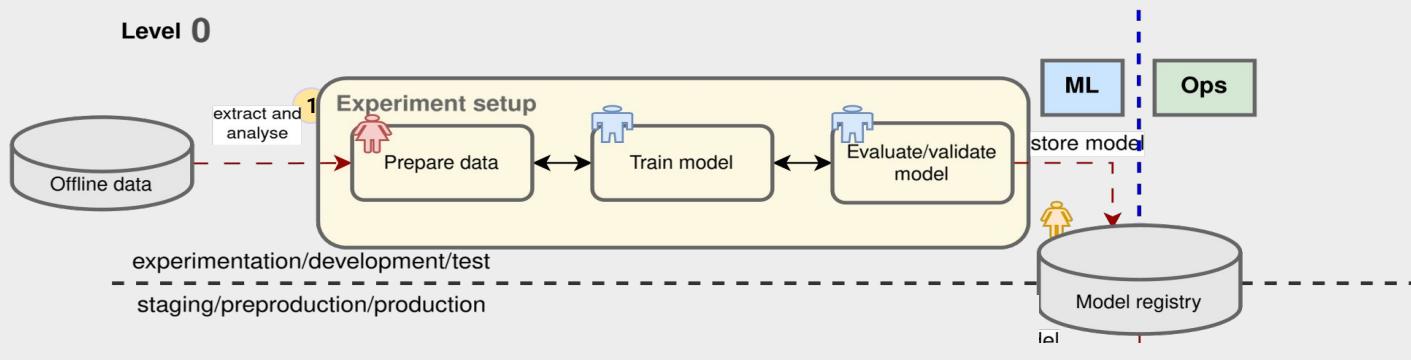
MLOps components

Adapted from Google

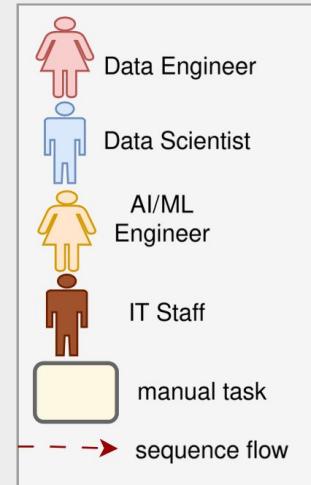
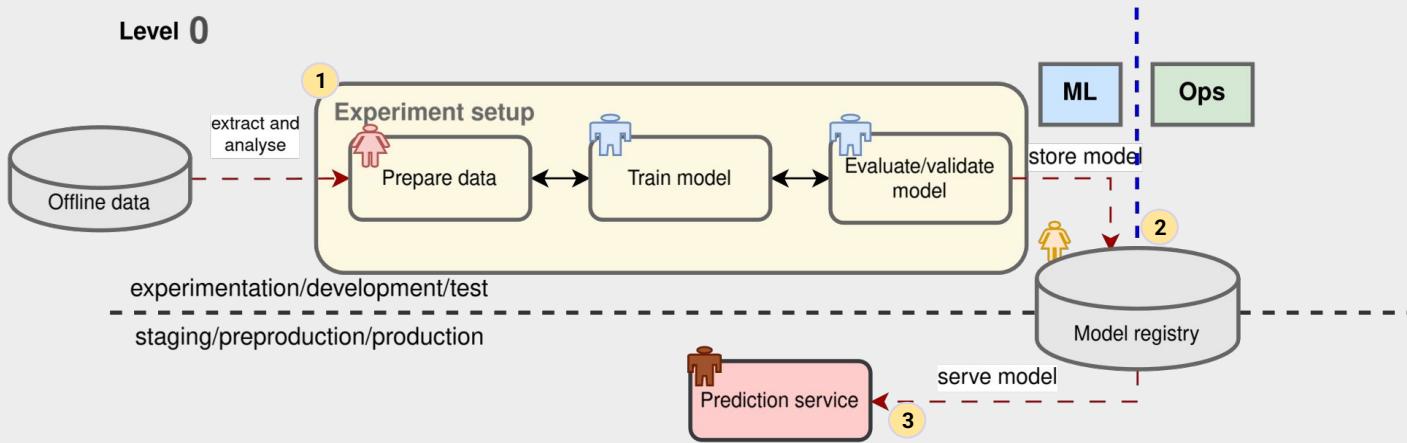
([MLOps: Continuous delivery and automation pipelines in machine learning](#))



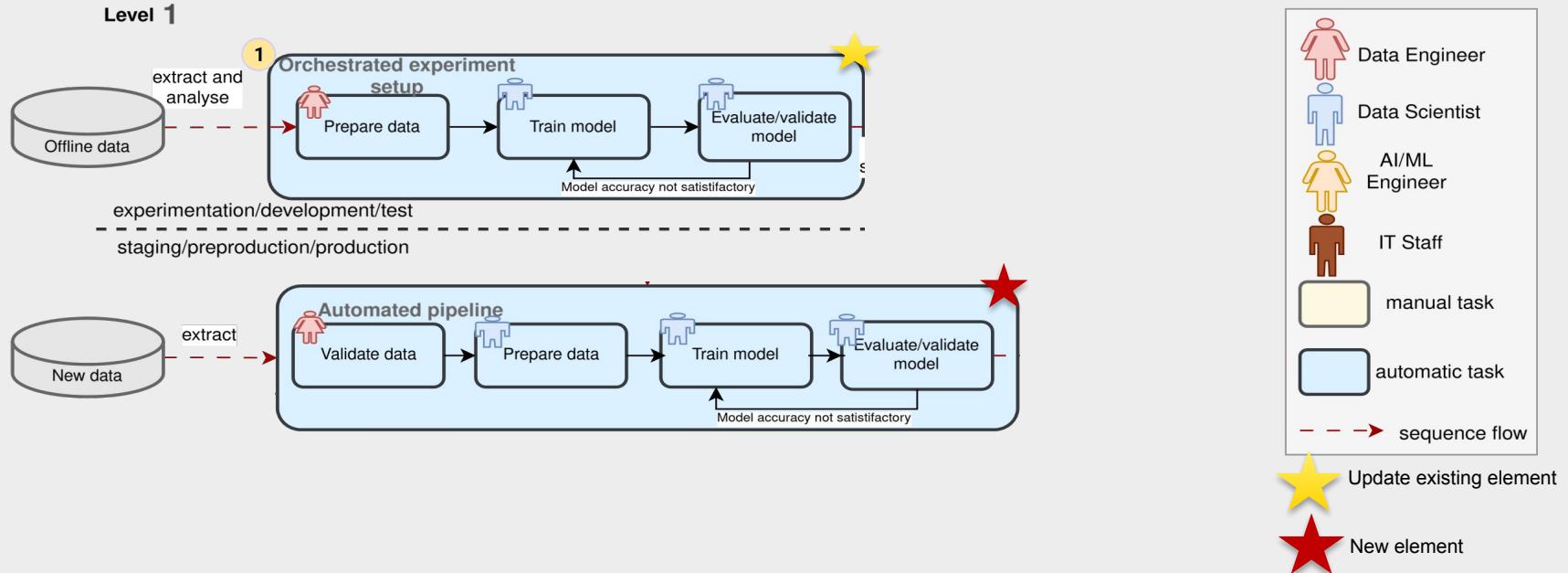
MLOps components



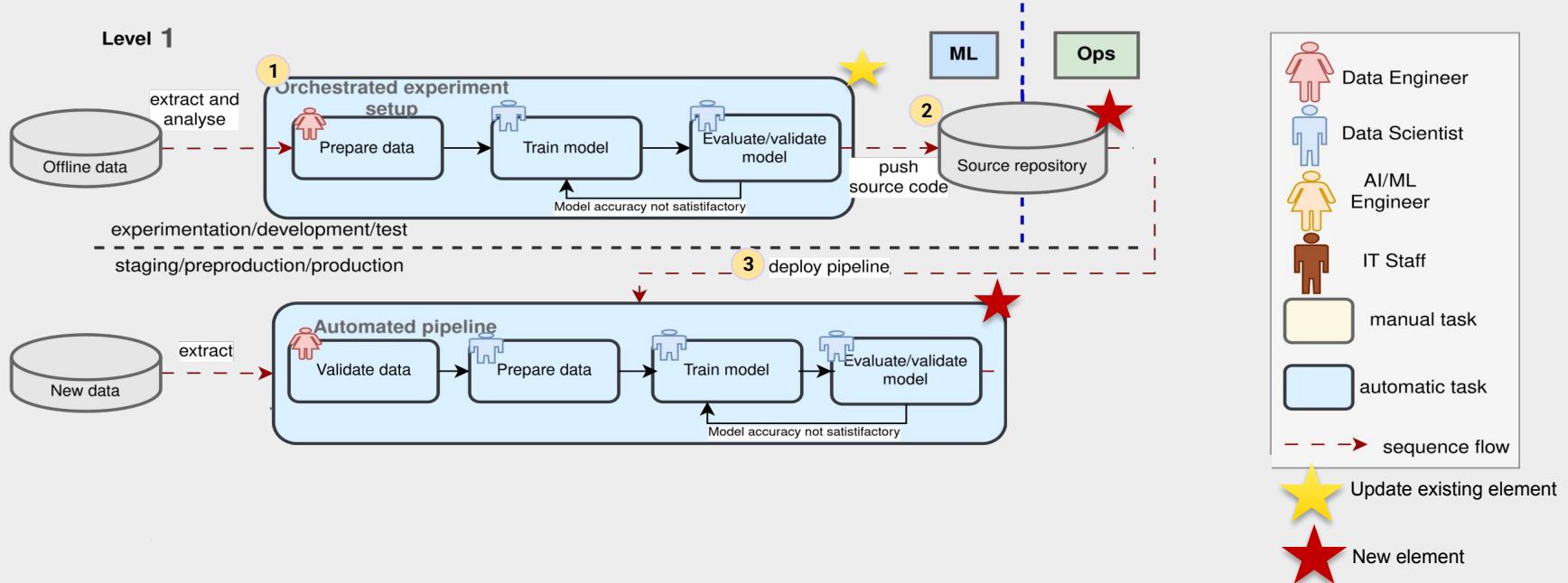
MLOps components



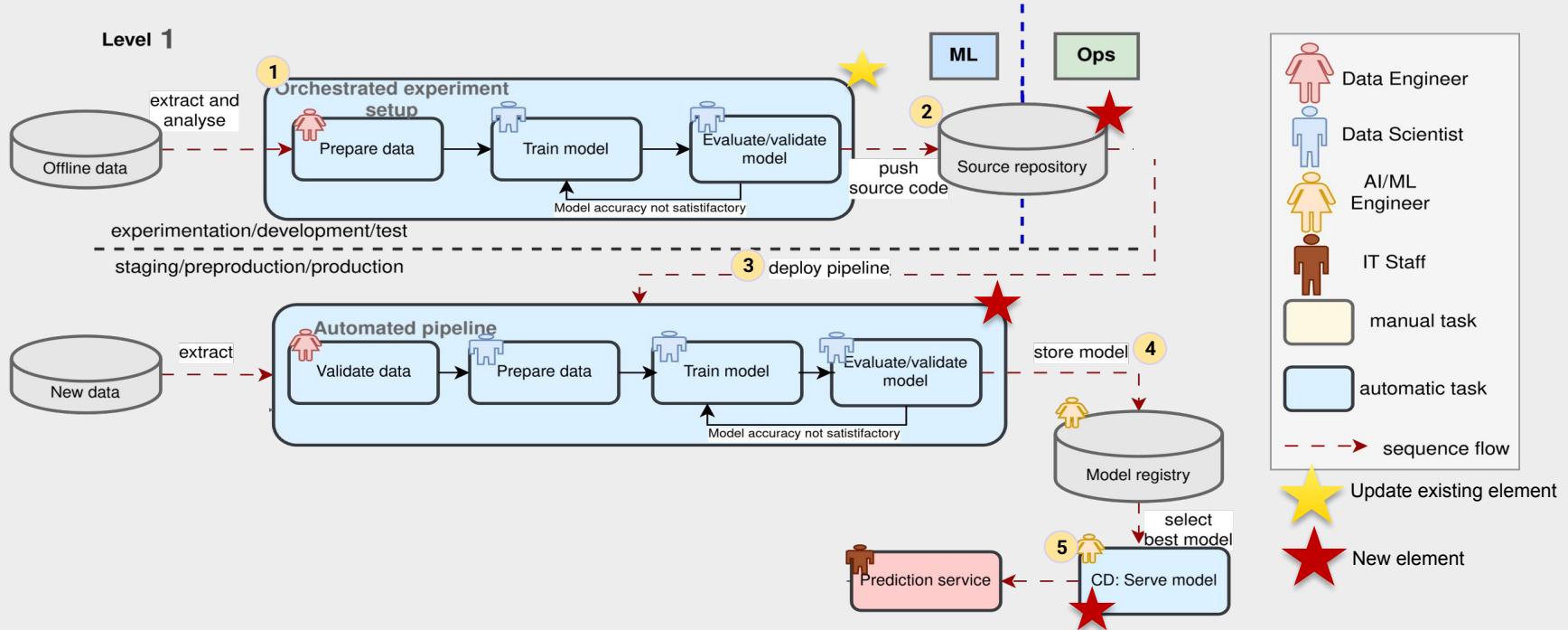
MLOps components



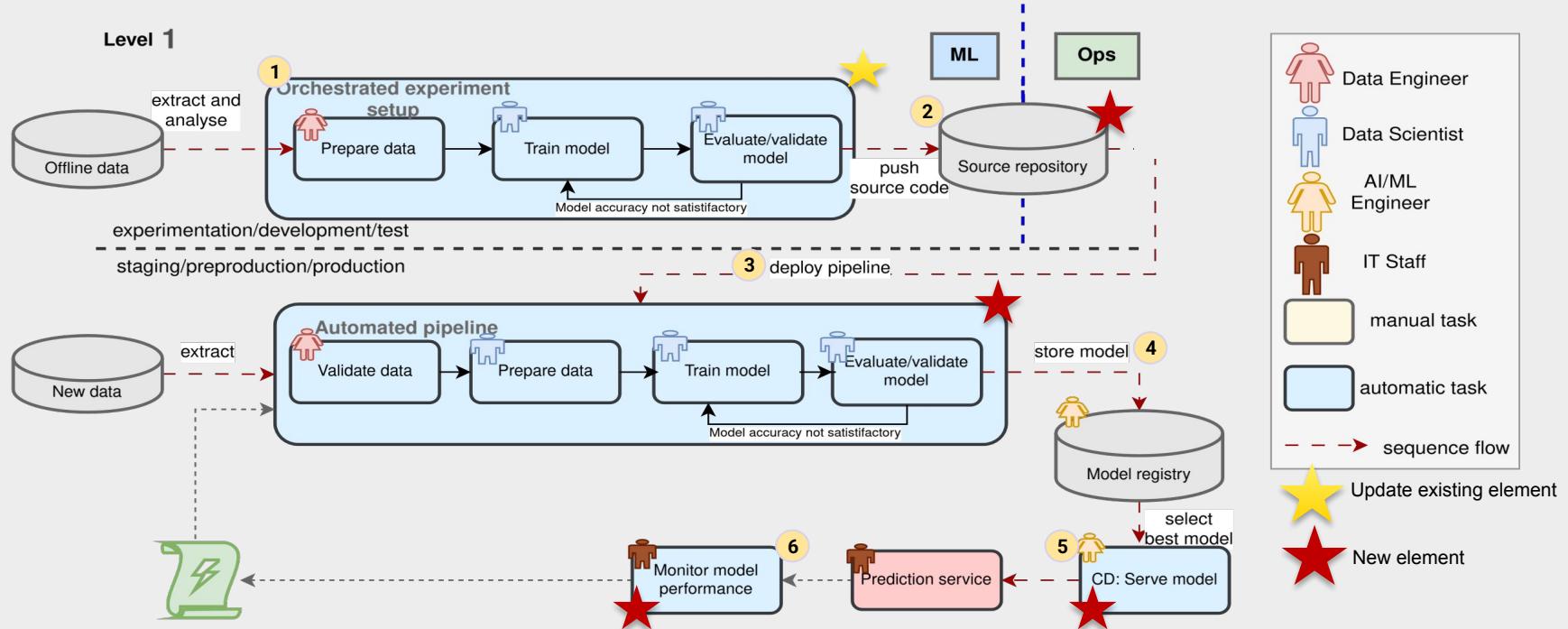
MLOps components

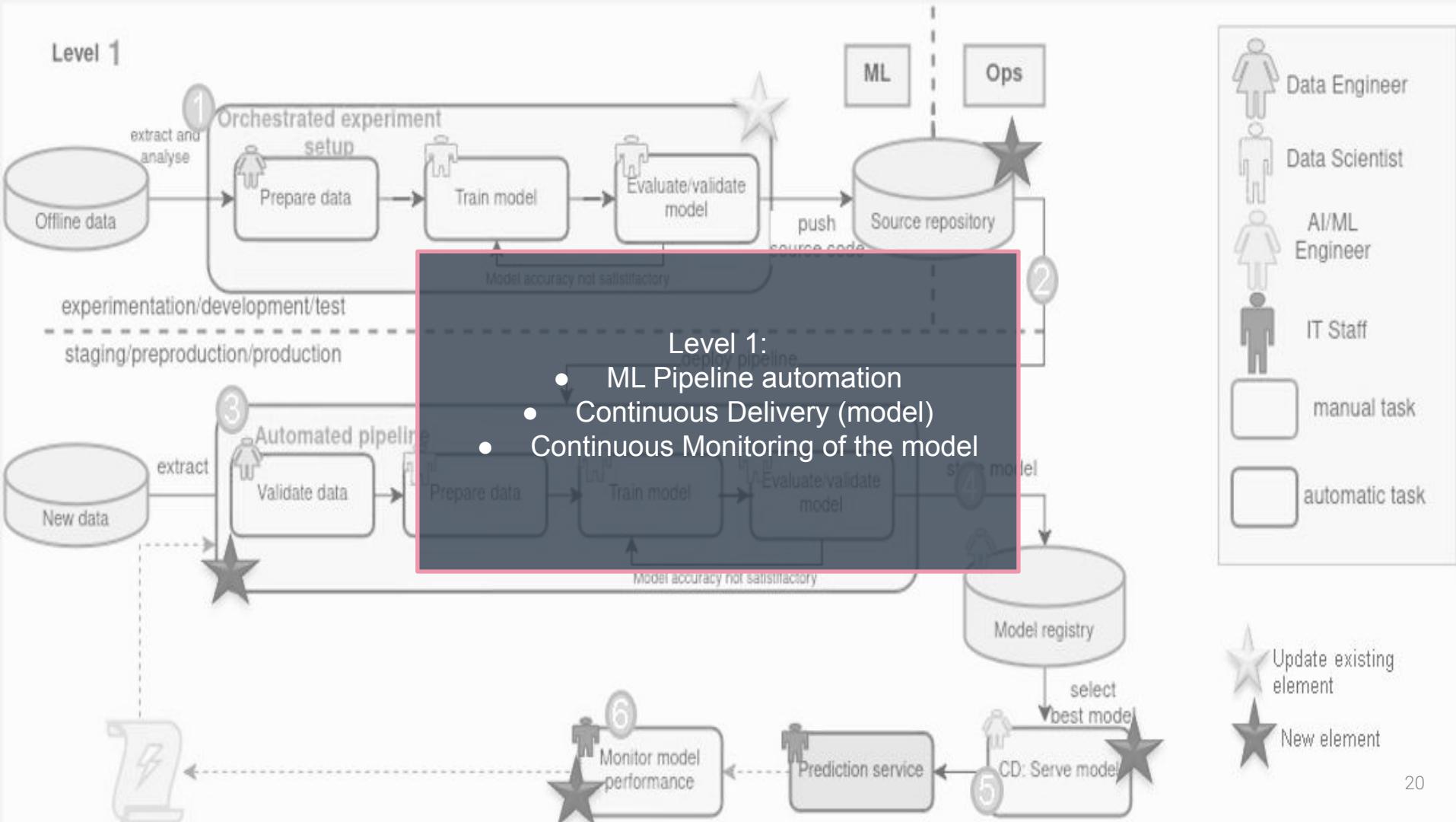


MLOps components

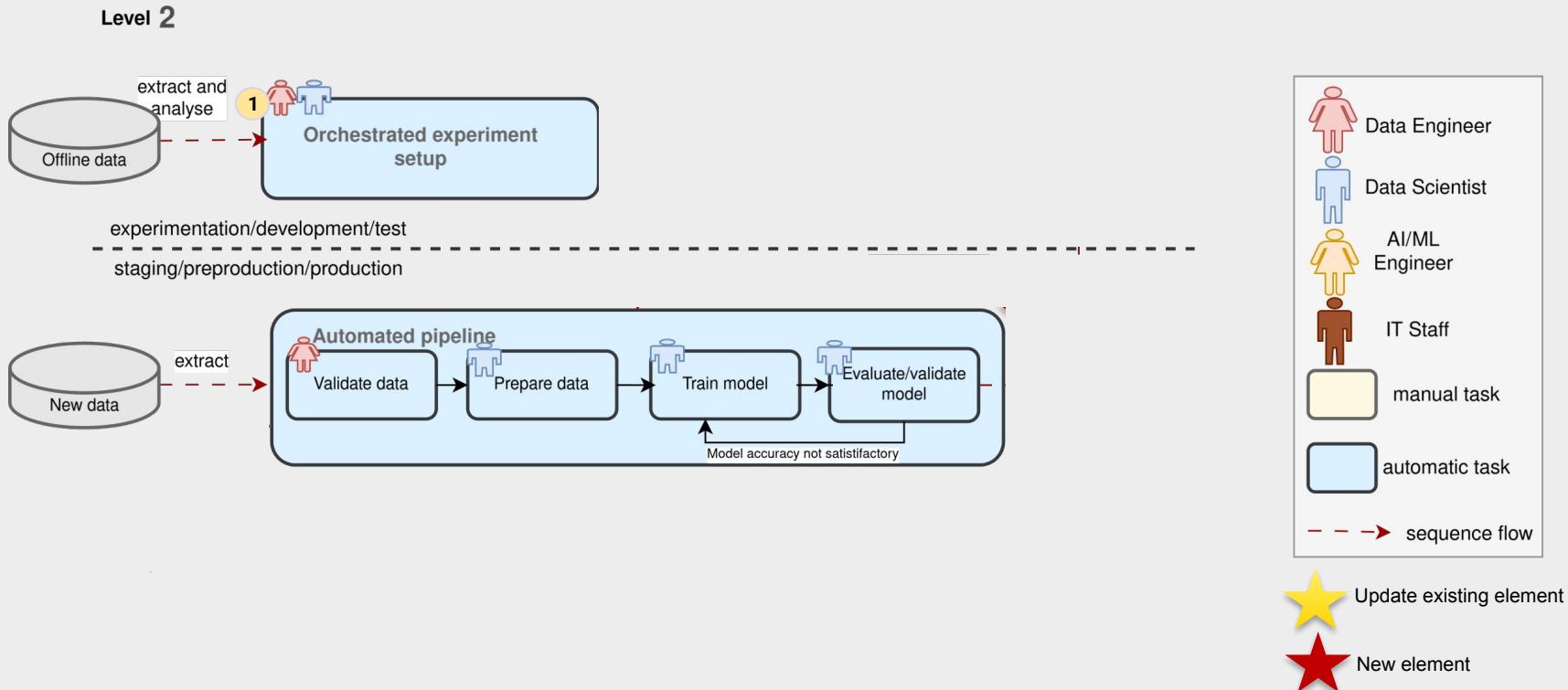


MLOps components

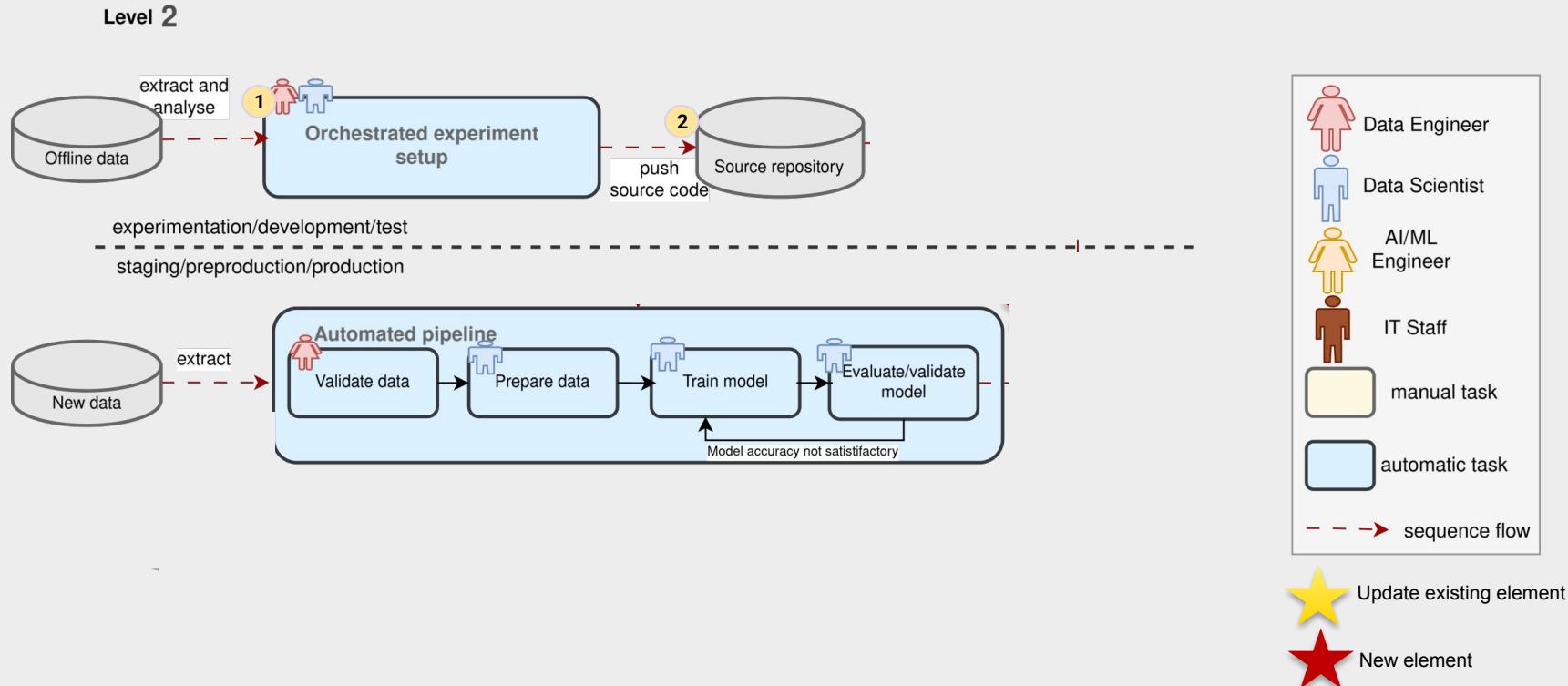




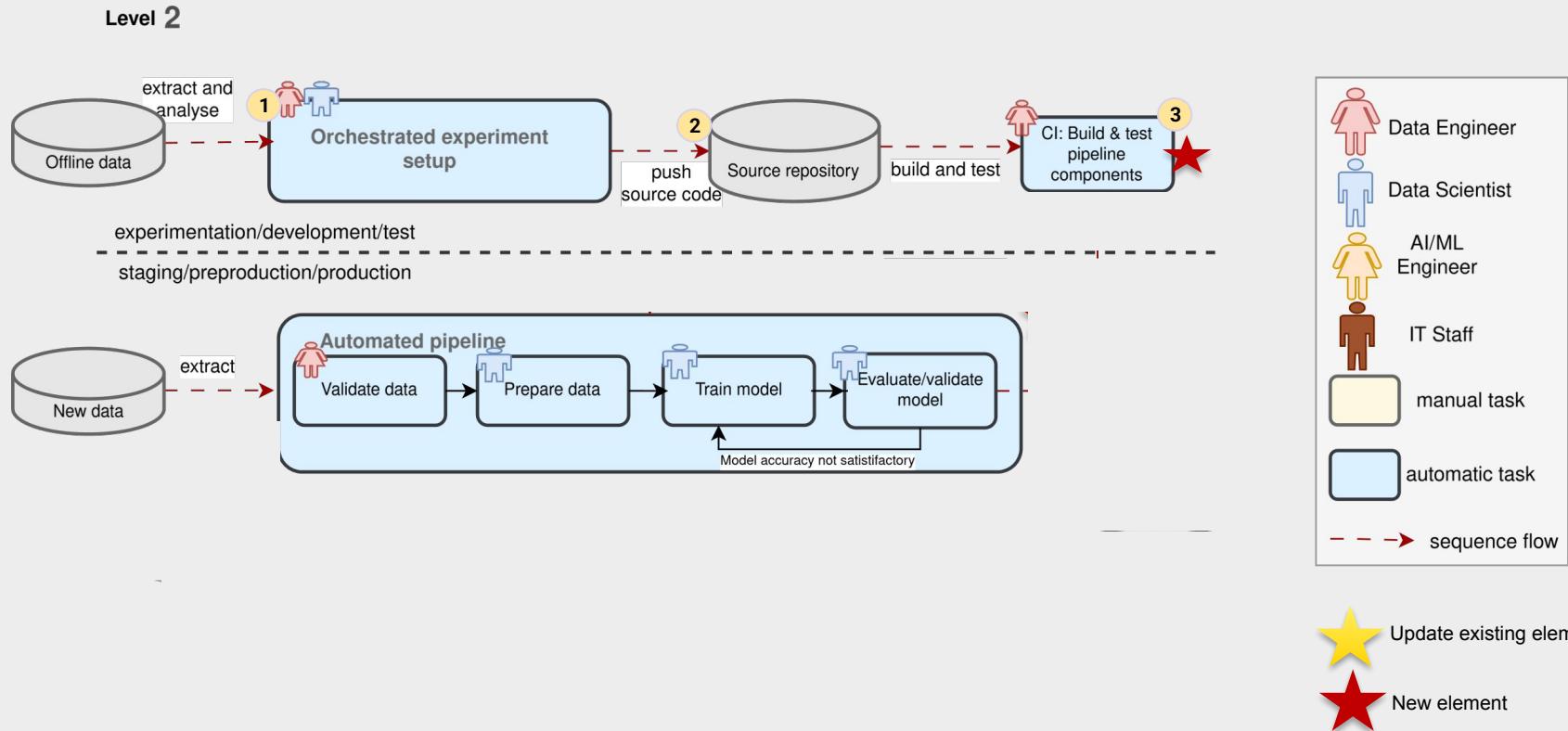
MLOps components



MLOps components

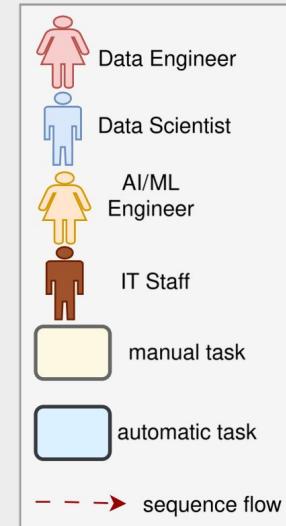
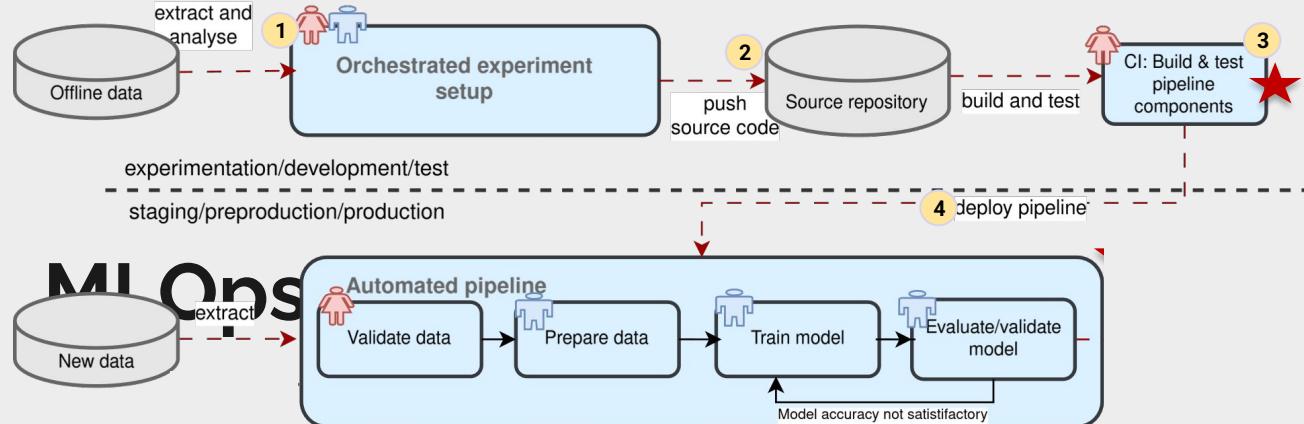


MLOps components



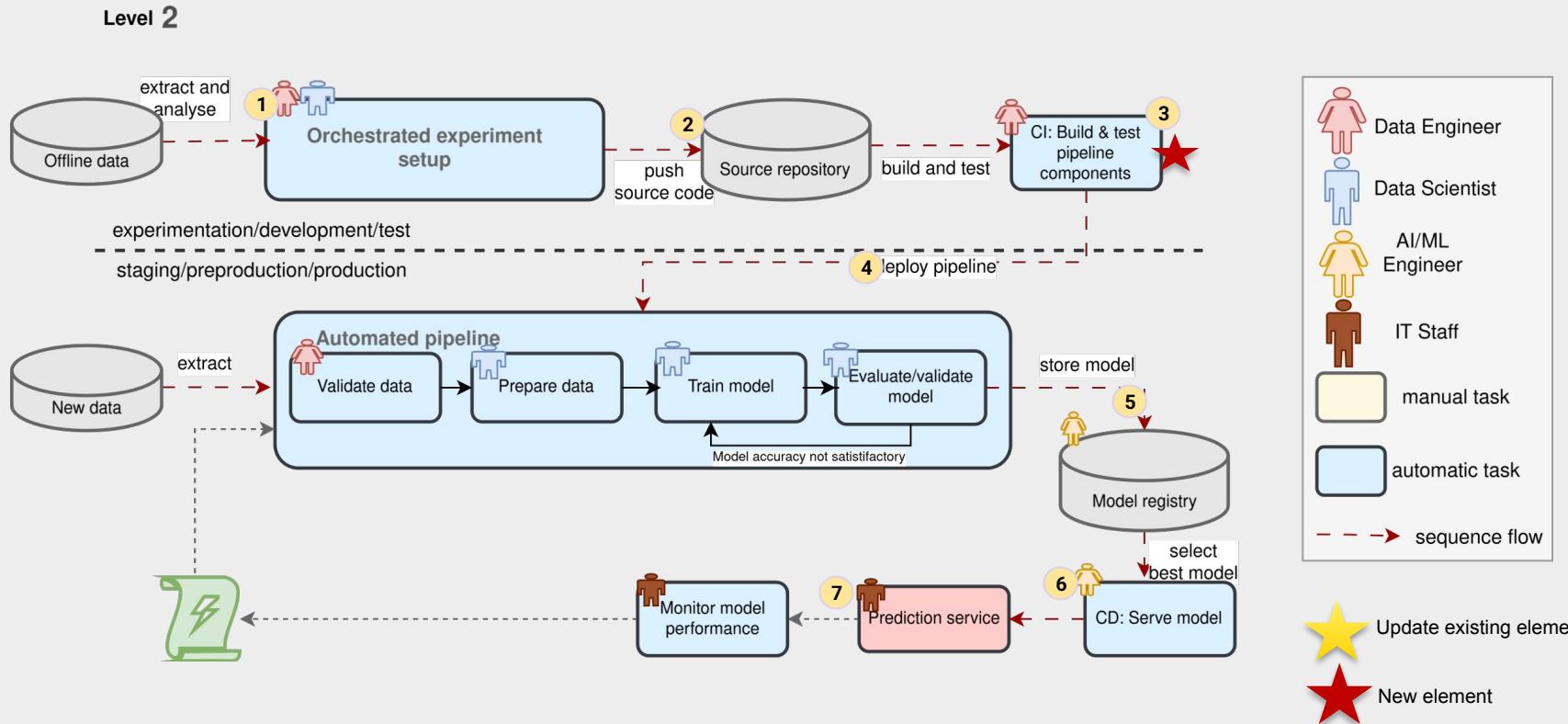
★ Update existing element
★ New element

Level 2

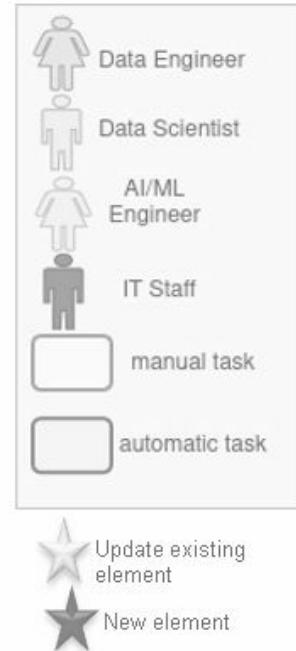
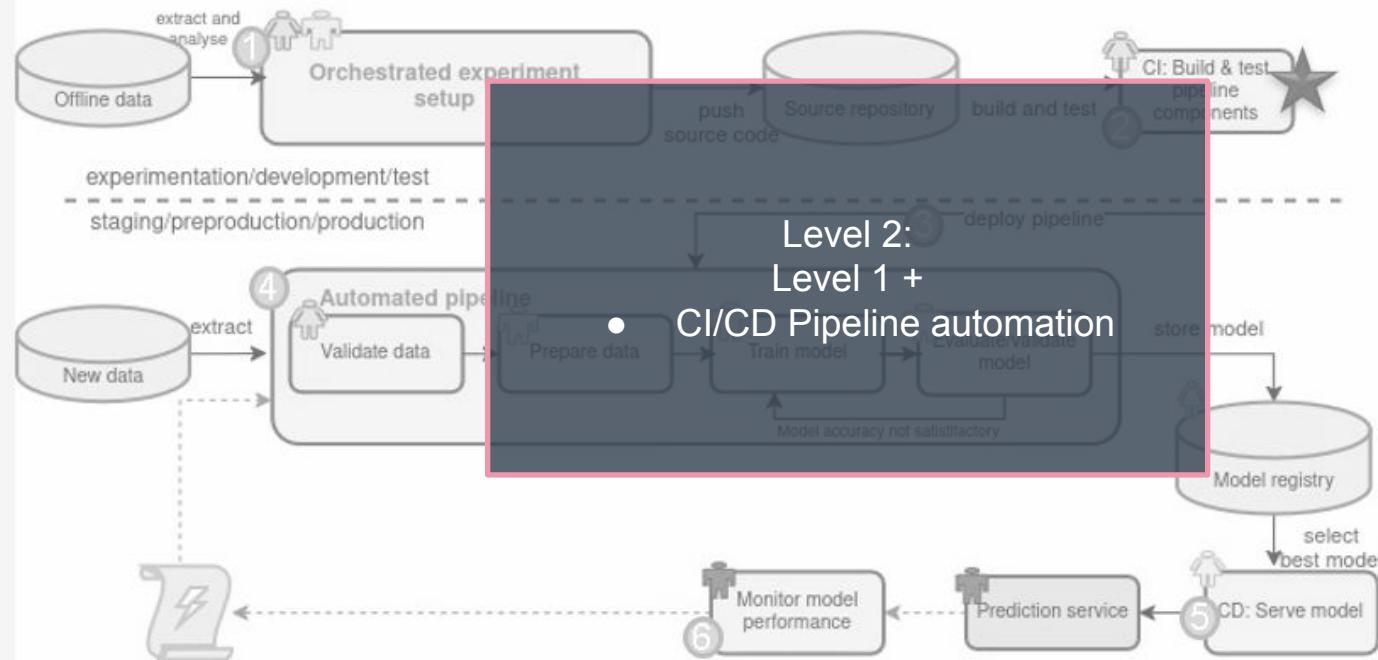


Update existing element

 New element



Level 2

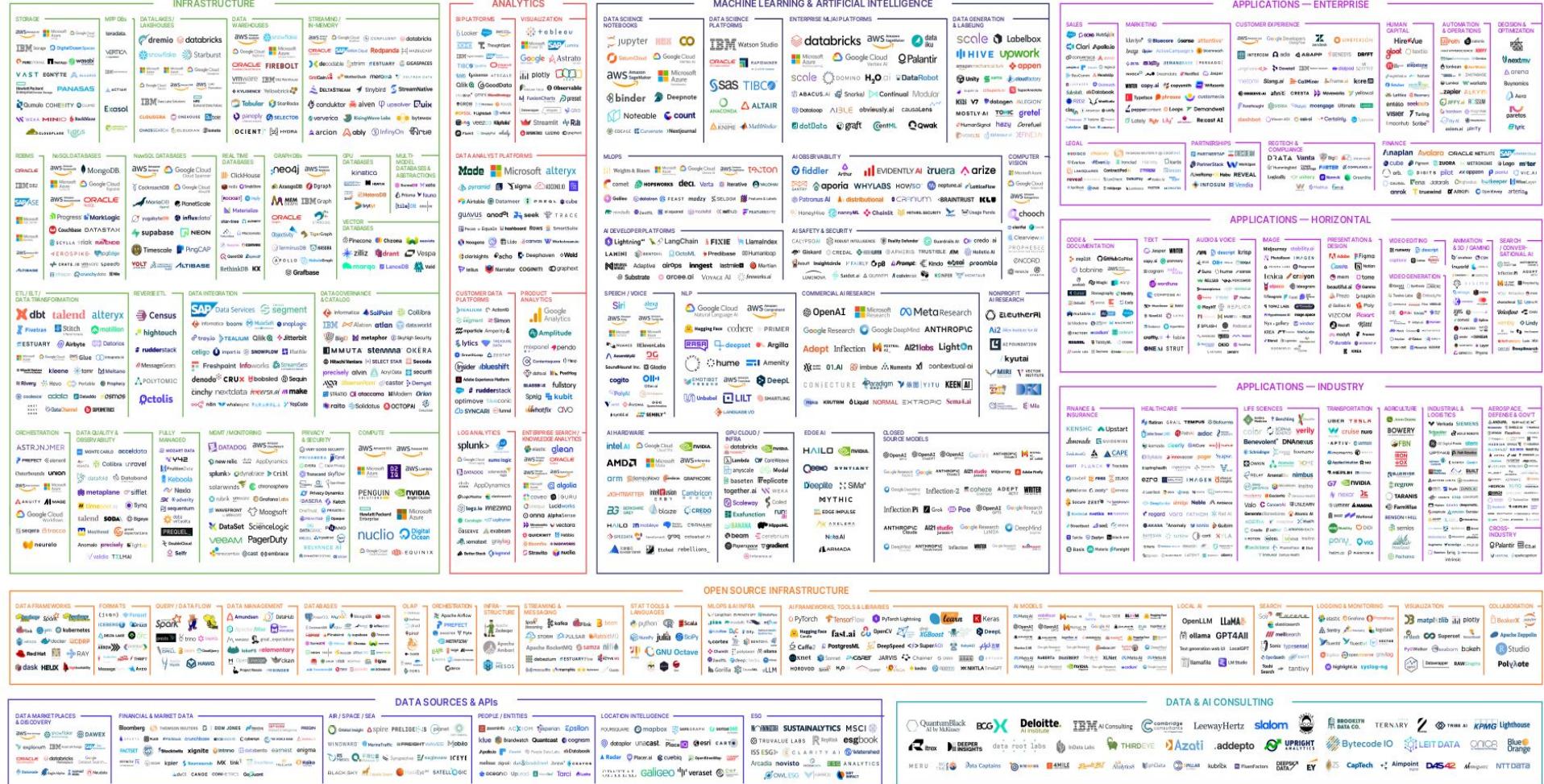


Funded by
the European Union

11 | 04 | 2025 – Lisana Berberi

| MLOps Platform/Tools |

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



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.

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	O	DT	CM	MTV	ETMS	DVM	MPM	Full Score	Partial Score
	GitHub Stars	Distributed Orchestration	Code Management	MDV Model Development					
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	✓✓	✓✓	✓✓	✓✓	✓	✓✓	✓✓	
MLeap	1.5 K	✓✓				✓✓	✓✓		✓ 30% 10%
MLRun	1.5 K	✓✓	✓✓	✓	✓✓	✓	✓✓	✓✓	✓✓ 80% 20%

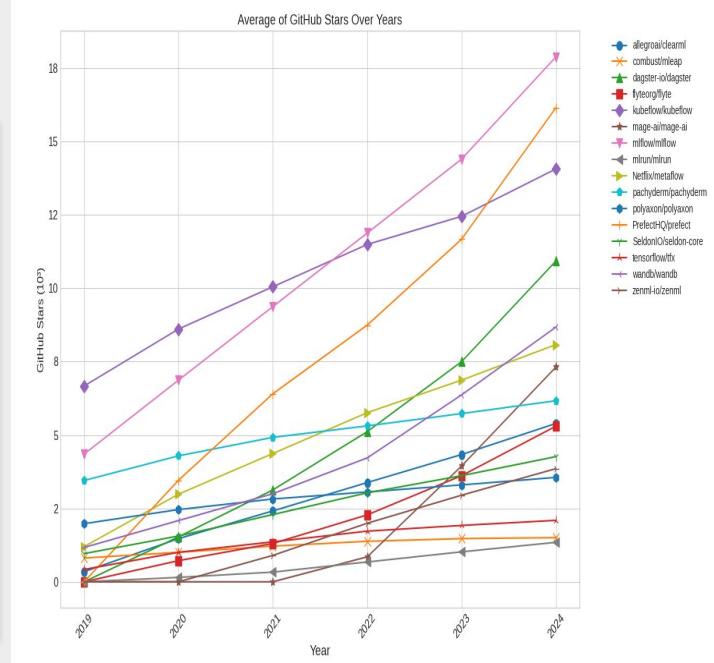
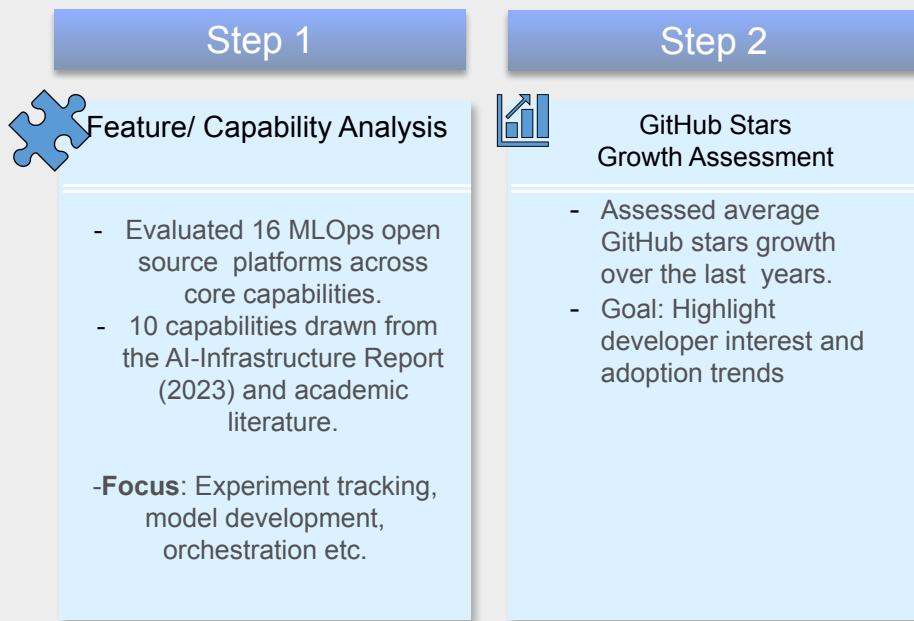
MLOps platforms/tools

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

Step 1	Step 2
 Feature/ Capability Analysis <ul style="list-style-type: none">- 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.	 GitHub Stars Growth Assessment <ul style="list-style-type: none">- Assessed average GitHub stars growth over the last years.- Goal: Highlight developer interest and adoption trends

MLOps platforms/tools

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



MLOps platforms/tools

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



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																																																																				
- Evaluation	Table 4 The weighted score for each product	<table><thead><tr><th>Product</th><th>Weight (w_i)</th><th>Feature-Score</th><th>Weighted-Score</th></tr></thead><tbody><tr><td>Kubeflow</td><td>8.89</td><td>7</td><td>62.21</td></tr><tr><td>WandB (W&B)</td><td>4.79</td><td>8</td><td>38.30</td></tr><tr><td>MLflow</td><td>10.00</td><td>3.5</td><td>35.00</td></tr><tr><td>Pachyderm</td><td>4.34</td><td>7.5</td><td>32.55</td></tr><tr><td>ClearML</td><td>2.94</td><td>10</td><td>29.38</td></tr><tr><td>Flyte</td><td>2.98</td><td>9</td><td>26.82</td></tr><tr><td>Polyaxon</td><td>2.65</td><td>9.5</td><td>25.22</td></tr><tr><td>ZenML</td><td>2.20</td><td>10</td><td>22.01</td></tr><tr><td>Dagster</td><td>6.13</td><td>3.5</td><td>21.44</td></tr><tr><td>Prefect</td><td>8.35</td><td>2.5</td><td>20.87</td></tr><tr><td>Mage</td><td>3.98</td><td>4.5</td><td>17.90</td></tr><tr><td>Metaflow</td><td>4.11</td><td>3.5</td><td>14.40</td></tr><tr><td>Seldon core</td><td>2.59</td><td>5.5</td><td>14.23</td></tr><tr><td>TFX</td><td>1.30</td><td>8</td><td>10.40</td></tr><tr><td>MLRun</td><td>1.00</td><td>9</td><td>9.00</td></tr><tr><td>MLeap</td><td>1.32</td><td>3.5</td><td>4.61</td></tr></tbody></table>	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	
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																																																																				

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																																																																			
- Focus	<p>Table 4 The weighted score for each product</p> <table><thead><tr><th>Product</th><th>Weight (w_i)</th><th>Feature-Score</th><th>Weighted-Score</th></tr></thead><tbody><tr><td>Kubeflow</td><td>8.89</td><td>7 Kubernetes bound</td><td>62.21</td></tr><tr><td>WandB (W&B)</td><td>4.79</td><td>8 limited features in the free version</td><td>38.30</td></tr><tr><td>MLflow</td><td>10.00</td><td>3.5</td><td>35.00</td></tr><tr><td>Pachyderm</td><td>4.34</td><td>7.5</td><td>32.55</td></tr><tr><td>ClearML</td><td>2.94</td><td>10</td><td>29.38</td></tr><tr><td>Flyte</td><td>2.98</td><td>9</td><td>26.82</td></tr><tr><td>Polyaxon</td><td>2.65</td><td>9.5</td><td>25.22</td></tr><tr><td>ZenML</td><td>2.20</td><td>10</td><td>22.01</td></tr><tr><td>Dagster</td><td>6.13</td><td>3.5</td><td>21.44</td></tr><tr><td>Prefect</td><td>8.35</td><td>2.5</td><td>20.87</td></tr><tr><td>Mage</td><td>3.98</td><td>4.5</td><td>17.90</td></tr><tr><td>Metaflow</td><td>4.11</td><td>3.5</td><td>14.40</td></tr><tr><td>Seldon core</td><td>2.59</td><td>5.5</td><td>14.23</td></tr><tr><td>TFX</td><td>1.30</td><td>8</td><td>10.40</td></tr><tr><td>MLRun</td><td>1.00</td><td>9</td><td>9.00</td></tr><tr><td>MLEap</td><td>1.32</td><td>3.5</td><td>4.61</td></tr></tbody></table>	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	
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																																																																			

MLOps platforms/tools

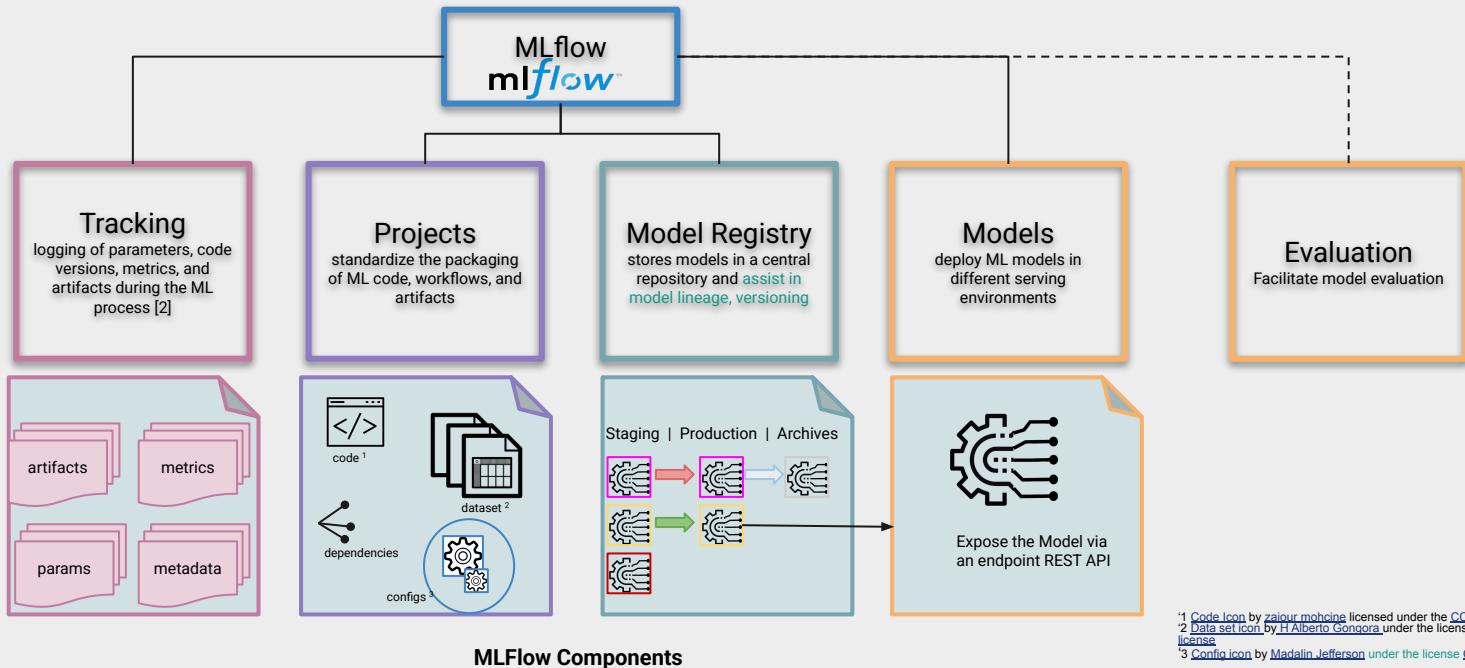
- Three-step evaluation framework [[L. Berberi et al. \(2025\)](#)]

	Step 1	Step 2	Step 3																																																																				
	 Feature/ Capability Analysis	 GitHub Stars Growth Assessment	 Weighted Scoring & Feature Extraction																																																																				
- Focus	Table 4 The weighted score for each product	<table><thead><tr><th>Product</th><th>Weight (w_i)</th><th>Feature-Score</th><th>Weighted-Score</th></tr></thead><tbody><tr><td>Kubeflow</td><td>8.89</td><td>7</td><td>62.21</td></tr><tr><td>WandB (W&B)</td><td>4.79</td><td>8</td><td>38.30</td></tr><tr style="outline: 2px solid red;"><td>MLflow</td><td>10.00</td><td>3.5</td><td>35.00</td></tr><tr><td>Pachyderm</td><td>4.34</td><td>7.5</td><td>32.55</td></tr><tr><td>ClearML</td><td>2.94</td><td>10</td><td>29.38</td></tr><tr><td>Flyte</td><td>2.98</td><td>9</td><td>26.82</td></tr><tr><td>Polyaxon</td><td>2.65</td><td>9.5</td><td>25.22</td></tr><tr><td>ZenML</td><td>2.20</td><td>10</td><td>22.01</td></tr><tr><td>Dagster</td><td>6.13</td><td>3.5</td><td>21.44</td></tr><tr><td>Prefect</td><td>8.35</td><td>2.5</td><td>20.87</td></tr><tr><td>Mage</td><td>3.98</td><td>4.5</td><td>17.90</td></tr><tr><td>Metaflow</td><td>4.11</td><td>3.5</td><td>14.40</td></tr><tr><td>Seldon core</td><td>2.59</td><td>5.5</td><td>14.23</td></tr><tr><td>TFX</td><td>1.30</td><td>8</td><td>10.40</td></tr><tr><td>MLRun</td><td>1.00</td><td>9</td><td>9.00</td></tr><tr><td>MLeap</td><td>1.32</td><td>3.5</td><td>4.61</td></tr></tbody></table>	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	
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																																																																				

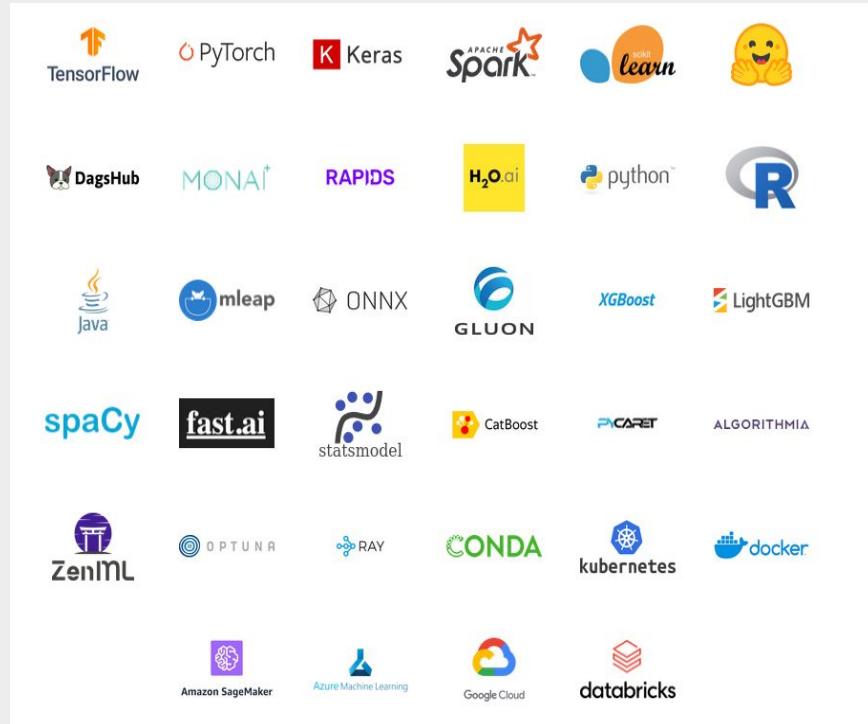
| MLflow |

Introduction to MLflow

- An open source platform for the machine learning lifecycle
- mlflow 3.6.0 (latest [release](#))



MLflow integrations and community support



[MLflow Integrations Logos](#) by [MLflow Project, a Series of LF Projects, LLC](#) licensed under the [CC BY 4.0 license](#)



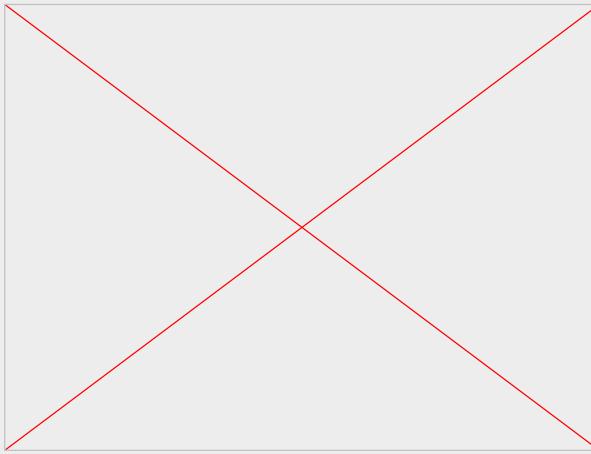
[MLflow community](#) by [MLflow Project, a Series of LF Projects, LLC](#) licensed under the [CC BY 4.0 license](#)

| Key takeaways |

Conclusions

- MLOps is critical for scaling machine learning beyond experimentation into production.
- MLOps maturity levels help assess and plan ML lifecycle automation.
- MLflow is a versatile and widely used tool that supports experiment tracking, model registry, model deployment

| MLflow experiment tracking and deployment demo |





GitHub repo link