MLOps platforms review: case study for AI4EOSC

Presentation · June 2023

DOI: 10.13140/RG.2.2.13326.33609

Lisana Berberi
Karlsruhe Institute of Technology

Jaime Céspedes Sisniega
Instituto de Física de Cantabria

MLOps platforms review: case study for AI4EOSC

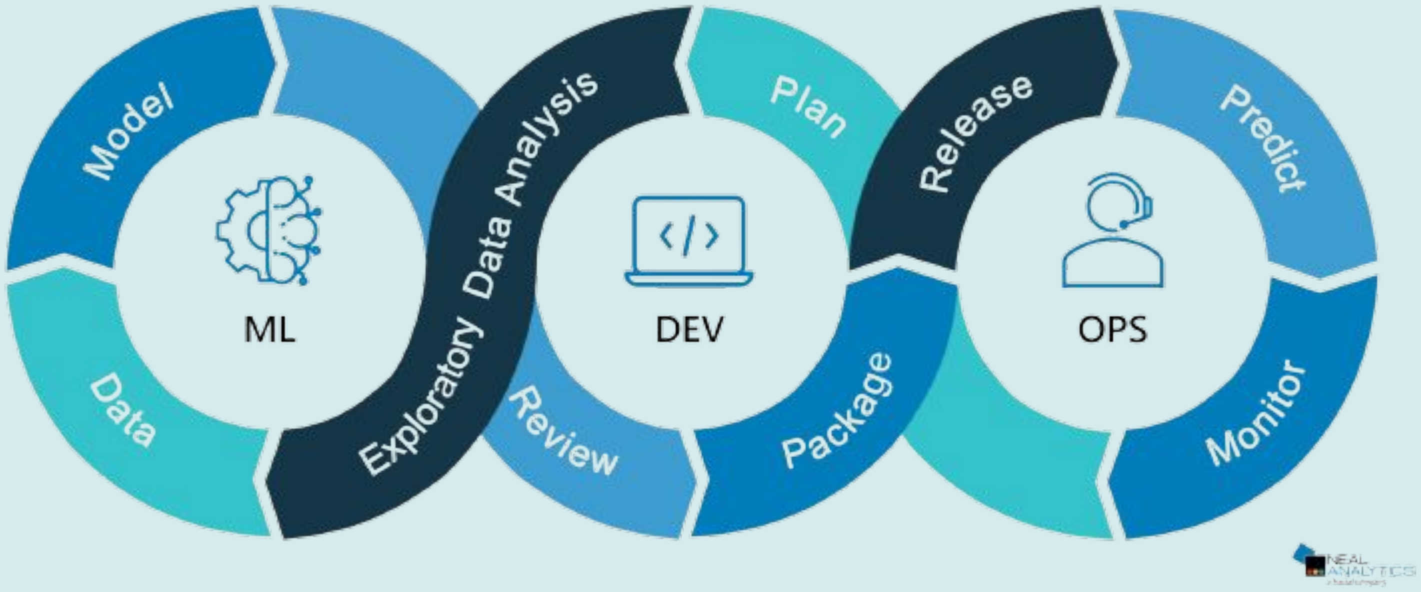
L. Berberi, V. Kozlov, J. Céspedes Sisniega, Á. López García

Co-funded by the European Union

Overview

- Introduction on MLOps
- AI4EOSC Project: Use case requirements
- Landscaping of MLOps open-source frameworks
- Conclusions

What is MLOps?



MLOps in a nutshell (source: Neal Analytics [1])

MLOps Roadmap-Maturity level

AI4 | eosc   Co-funded by the European Union

0
No Ops

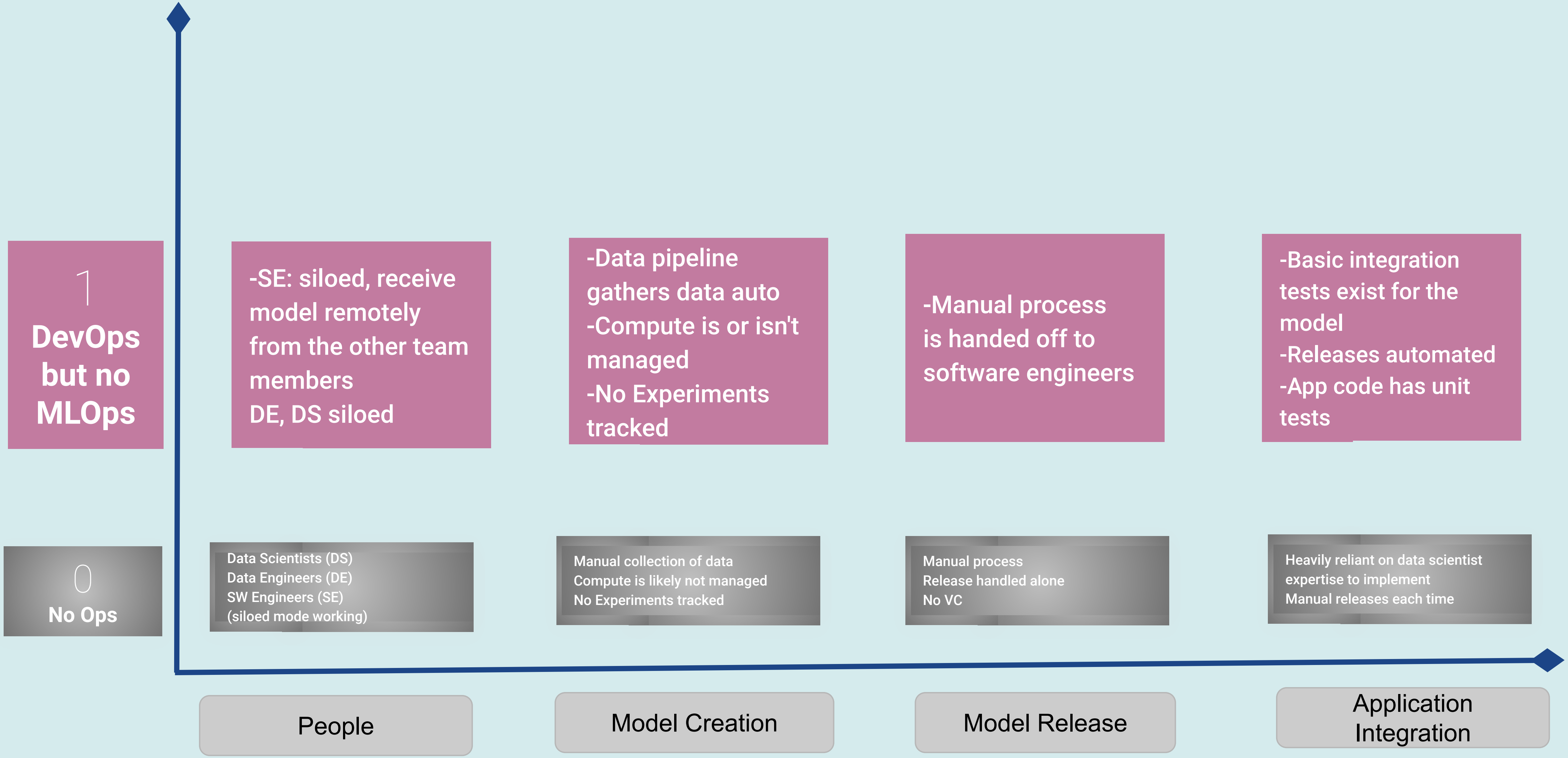
People	Model Creation	Model Release	Application Integration
-Data Scientists (DS) -Data Engineers (DE) -SW Engineers (SE) (siloed mode working)	-Manual collection of data -Compute is likely not managed -No Experiments tracked	-Manual process -Release handled alone -No VC	-Heavily reliant on data scientist expertise to implement -Manual releases each time

MLOps Roadmap (adopted from [2])

MLOps Roadmap-Maturity level

AI4 | eosc

Co-funded by the European Union



MLOps Roadmap (adopted from [2])

MLOps Roadmap-Maturity level

AI4 | eosc
Co-funded by the European Union

	People	Model Creation	Model Release	Application Integration
2 **Automated training**	-DS<->DE (convert experimentation code into repeatable scripts/jobs) -SE siloed with rest of team	-Data pipeline gathers data auto -Compute is managed -Experiments tracked Training code + Res.Models are VC	-Manual release -Scoring script is VC with tests -Release managed by SE team	-Basic integration tests exist for the model -Heavily reliant on data scientist expertise to implement -App code has unit tests
1 **DevOps but no MLOps**	SE: siloed, receive model remotely from the other team members DE, DS siloed	Data pipeline gathers data auto Compute is or isn't managed No Experiments tracked	Manual process Is handed off to software engineers	Basic integration tests exist for the model Releases automated App code has unit tests

MLOps Roadmap (adopted from [2])

MLOps Roadmap-Maturity level

AI4 | eosc

Co-funded by the European Union

	People	Model Creation	Model Release	Application Integration
3 **Automated MD**	-DS<->DE<->SE (to automate model integration into application code)	Same as Level 2	-Auto release -Scoring script is VC with tests -Release managed by CI/CD pipeline	-Unit and integration tests for each model release -Less reliant on DS expertise to implement model
2 **Automated training**	DS<->DE (convert experimentation code into repeatable scripts/jobs) SE siloed with rest of team	Data pipeline gathers data auto Compute is managed Experiments tracked Training code + Res.Models are VC	Manual release Scoring script is VC with tests Release managed by SE team	Basic integration tests exist for the model Heavily reliant on data scientist expertise to implement App code has unit tests

MLOps Roadmap (adopted from [2])

MLOps Roadmap-Maturity level

AI4 | eosc
Co-funded by the European Union

	People	Model Creation	Model Release	Application Integration
4 **Automated Ops**	-DS<->DE<->SE (to automate model integration into application code) Implementing post-deployment metrics gathering	-Retraining triggered auto based on production metrics + all aspects from Level 3	-Same as Level 3	-Same as Level 3
3 **Automated MD**	DS<->DE<->SE (to automate model integration into application code)	Same as Level 2	Auto release Scoring script is VC with tests Release managed by CI/CD pipeline	Unit and integration tests for each model release Less reliant on DS expertise to implement model

MLOps Roadmap (adopted from [2])

MLOps Roadmap-Maturity level

AI4 | eosc

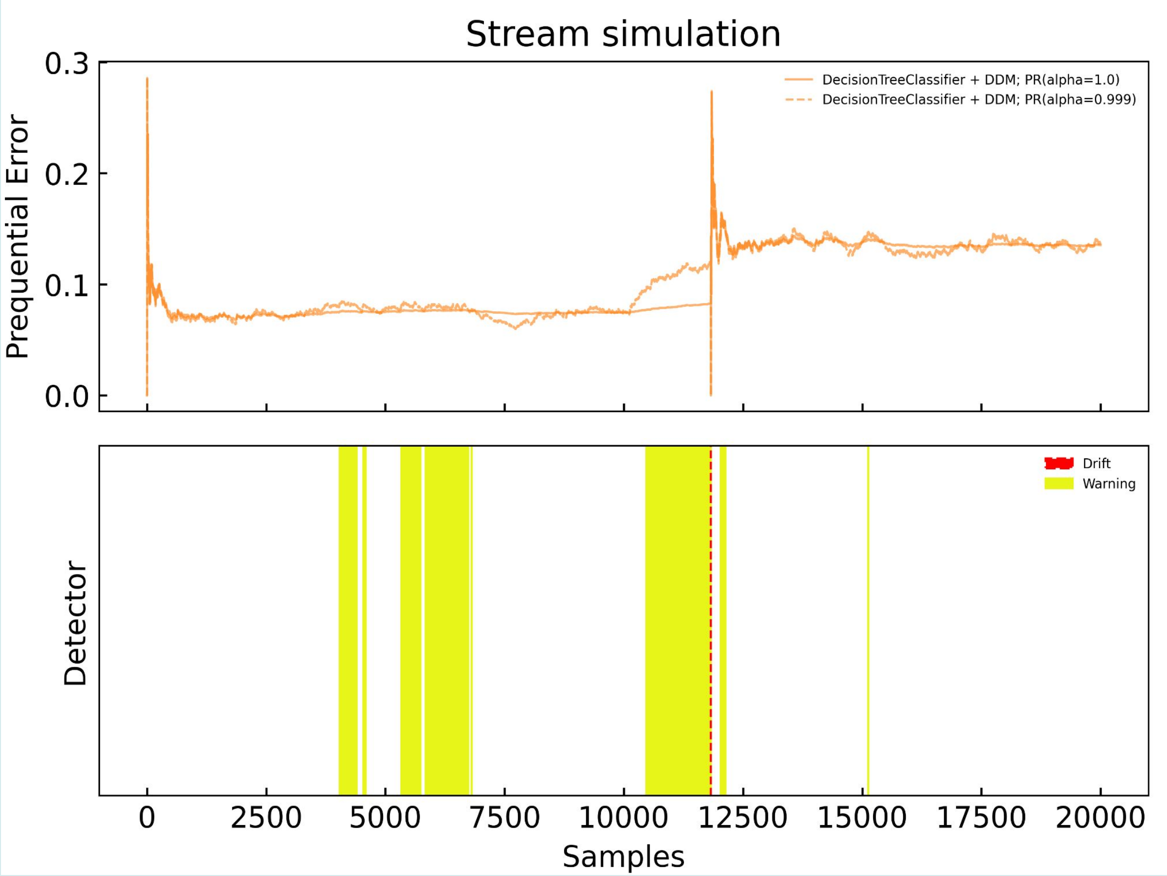
Co-funded by the European Union

Maturity level	People	Model Creation	Model Release	Application Integration
4 Automated Ops	DS<->DE<->SE (to automate model integration into application code) Implementing post-deployment metrics gathering	Retraining triggered auto based on production metrics + all aspects from Level 3	Same as Level 3	Same as Level 3
3 Automated MD	DS<->DE<->SE (to automate model integration into application code)	Same as Level 2	Auto release Scoring script is VC with tests Release managed by CI/CD pipeline	Unit and integration tests for each model release Less reliant on DS expertise to implement model
2 Automated training	DS<->DE (convert experimentation code into repeatable scripts/jobs) SE siloed with rest of team	Data pipeline gathers data auto Compute is managed Experiments tracked Training code + Res.Models are VC	Manual release Scoring script is VC with tests Release managed by SE team	Basic integration tests exist for the model Heavily reliant on data scientist expertise to implement App code has unit tests
1 DevOps but no MLOps	SE: siloed, receive model remotely from the other team members DE, DS siloed	Data pipeline gathers data auto Compute is or isn't managed No Experiments tracked	Manual process Is handed off to software engineers	Basic integration tests exist for the model Releases automated App code has unit tests
0 No Ops	Data Scientists (DS) Data Engineers (DE) SW Engineers (SE) (siloed mode working)	Manual collection of data Compute is likely not managed No Experiments tracked	Manual process Release handled alone No VC	Heavily reliant on data scientist expertise to implement Manual releases each time

MLOps Roadmap (adopted from [2])

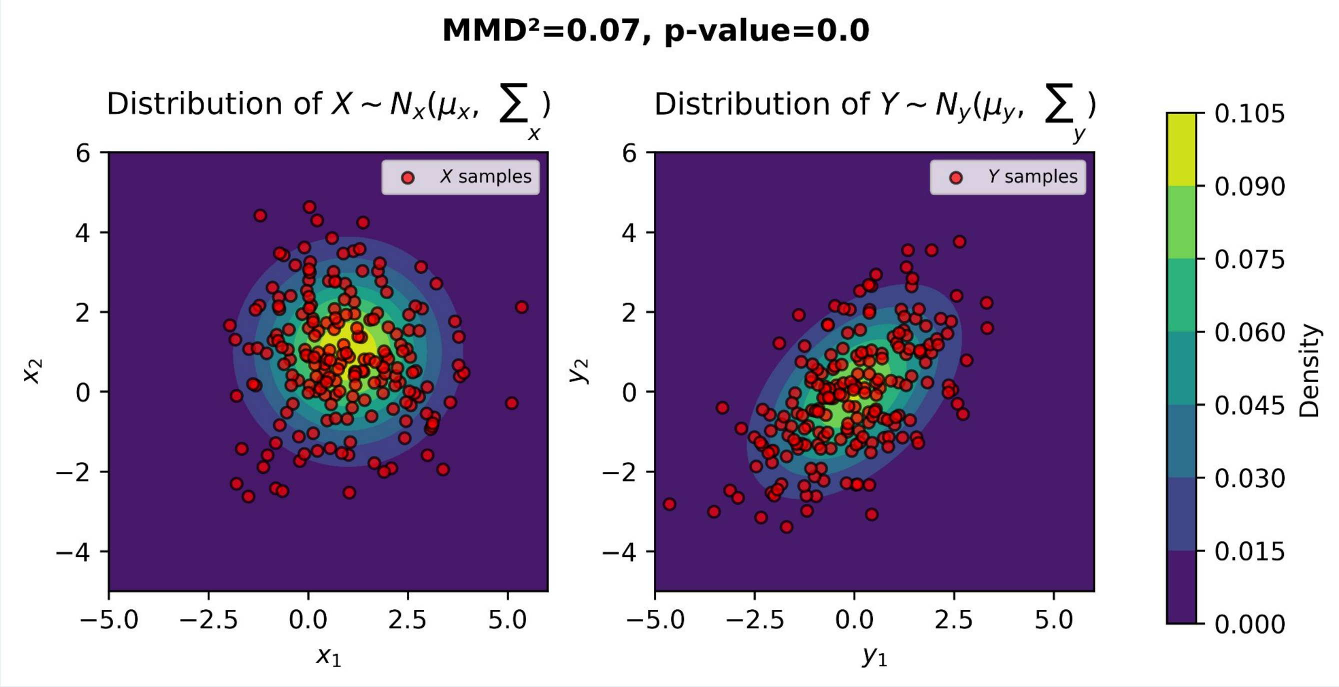
AI4 eosc
Co-funded by the European Union

Drift Detection



Concept drift detector ([7])

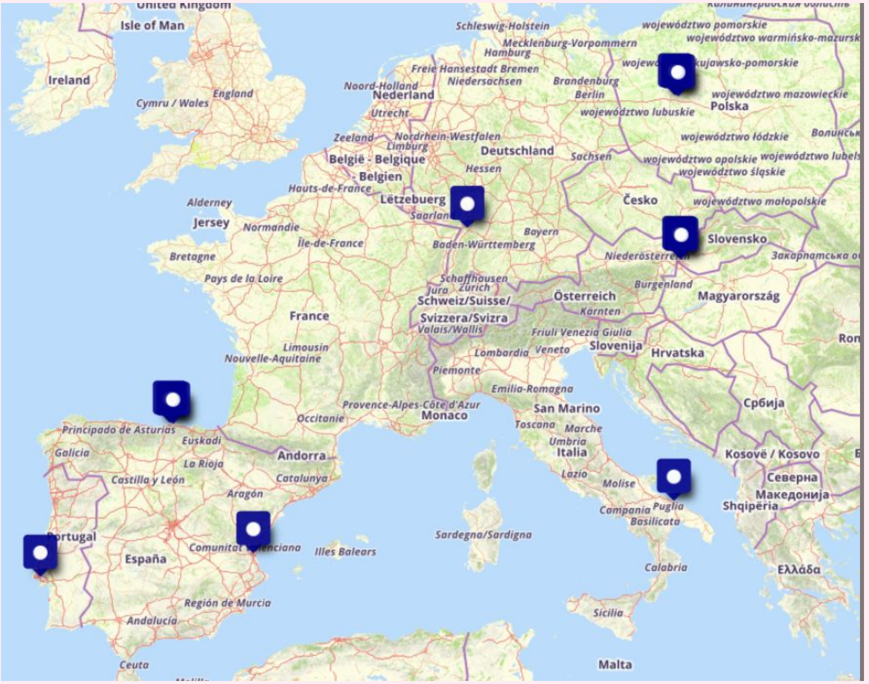
FROUROS



Data drift detector ([7])

AI4 eosc

Co-funded by the European Union



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

KIT
Karlsruher Institut für Technologie

LIP

ÚSTAV INFORMATIKY
SLOVENSKÁ AKADÉMIA VIED

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

INFN
Istituto Nazionale di Fisica Nucleare

PSNC

WODR

MicroStep-MIS

predictia

AI4EOSC

Artificial Intelligence for the #EOSC

- Evolution of the DEEP Hybrid DataCloud platform
- HORIZON-INFRA-2021-EOSC-01-04 call
- Runs September 1st 2022 – August 2025 (36 months)
- 7 academic partners
  + 2 SME
  + 1 non-profit organization

Advanced features for distributed, federated, composite learning, metadata provenance, MLOps, event-driven data processing, and provision of AI/ML/DL services

AI4 eosc Co-funded by the European Union

AI4EOSC Use Cases

UC1-Agrometeorology

AI product
Forecasting system

-**Problem solving**: Early warnings for farmers before approaching thunderstorms using AI techniques
-**Target users:** Farmers, public adm. etc.

UC2-Integrated Plant Protection

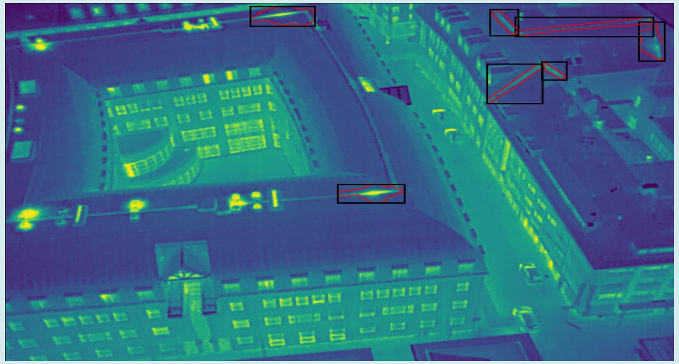
AI product
Recognizing plant diseases

-**Problem solving:** Reinforce the quality and quantity of food produced.
-**Target users**: Farmers, public administration, local governments etc.

UC3-Automated Thermography

AI product
Detection of thermal hotspots caused by thermal bridges and common urban features.

-**Problem solving:** Identifying energy losses to mitigate their effects and enable higher system efficiency.
-**Target users:** Urban planners, district heating network operators etc.



AI4EOSC Use Case Requirements

UC1.Req02/UC2.Req09/UC3.Req07- Organize and track all training experiments	Level 2
UC1:Req13- Monitor the model and concept drift for the inference	Level 3
UC2.Req02- Continuously monitor recently acquired input data sets	Level 3
UC2.Req15- Access the model validation results	Level 3
UC3.Req11- Monitor the model performance in production	Level 3

Co-funded by the European Union

Landscaping MLOps Products

Co-funded by the European Union

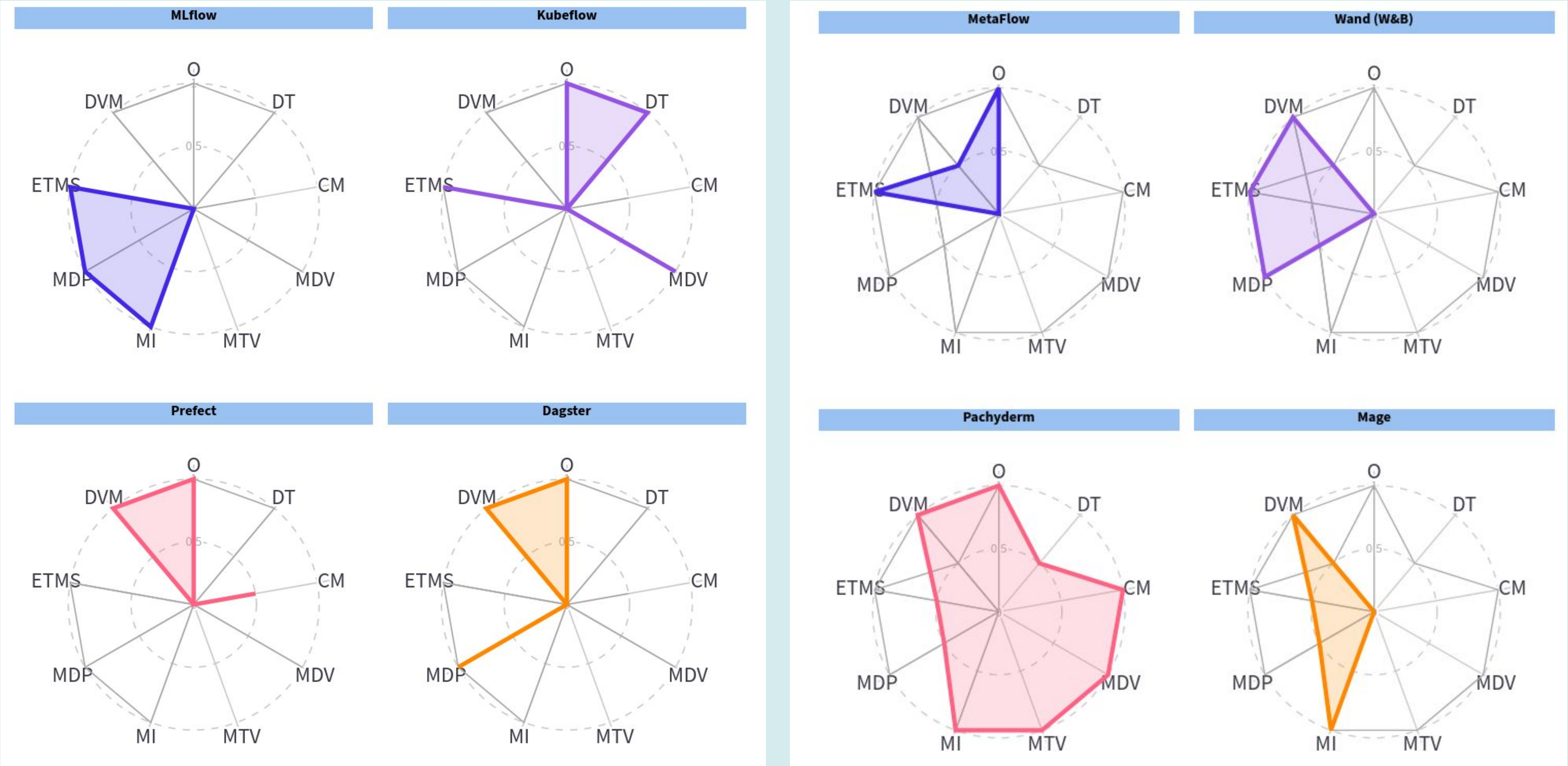
Product (platform)	GitHub Stars	O Orchestration	DT Distributed Training	CM Code Management	MDV Model Development	MTV Model Testing/Validation	MI Model Inference	MDP Model Deployment	ETMS Experiment Tracking and Metadata Store	DVM Data Versioning and Management
MLflow	14.6 K						✓✓	✓✓	✓✓	
Kubeflow	12.6 K	✓✓	✓✓		✓✓				✓✓	
Prefect	12.1 K	✓✓		✓						✓✓
Dagster	7.6 K	✓✓						✓✓		✓✓
MetaFlow	6.7 K	✓✓							✓✓	✓
Wand (W&B)	6.2 K							✓✓	✓✓	✓✓
Pachyderm	5.9 K	✓✓	✓	✓✓	✓✓	✓✓	✓✓	✓	✓	✓✓
Mage	4.8 K						✓✓	✓	✓	✓✓
ClearML	4.5 K	✓✓	✓✓	✓✓	✓✓	✓	✓✓	✓✓	✓✓	✓✓
Seldon core	3.8 K	✓✓				✓✓	✓✓	✓✓	✓	
Polyaxon	3.3 K	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓
Flyte	3.5 K	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓	✓✓
ZenML	2.9 K	✓✓		✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
TFX	2.0 K	✓✓	✓✓	✓✓	✓✓	✓	✓✓	✓✓	✓	✓✓
MLeap	1.5 K	✓✓						✓✓		
MLRun	983	✓✓	✓✓	✓	✓✓	✓	✓✓	✓✓	✓✓	✓✓

Notable Open Source MLOps Platforms

Landscaping: MLOps Frameworks

AI4 eosc

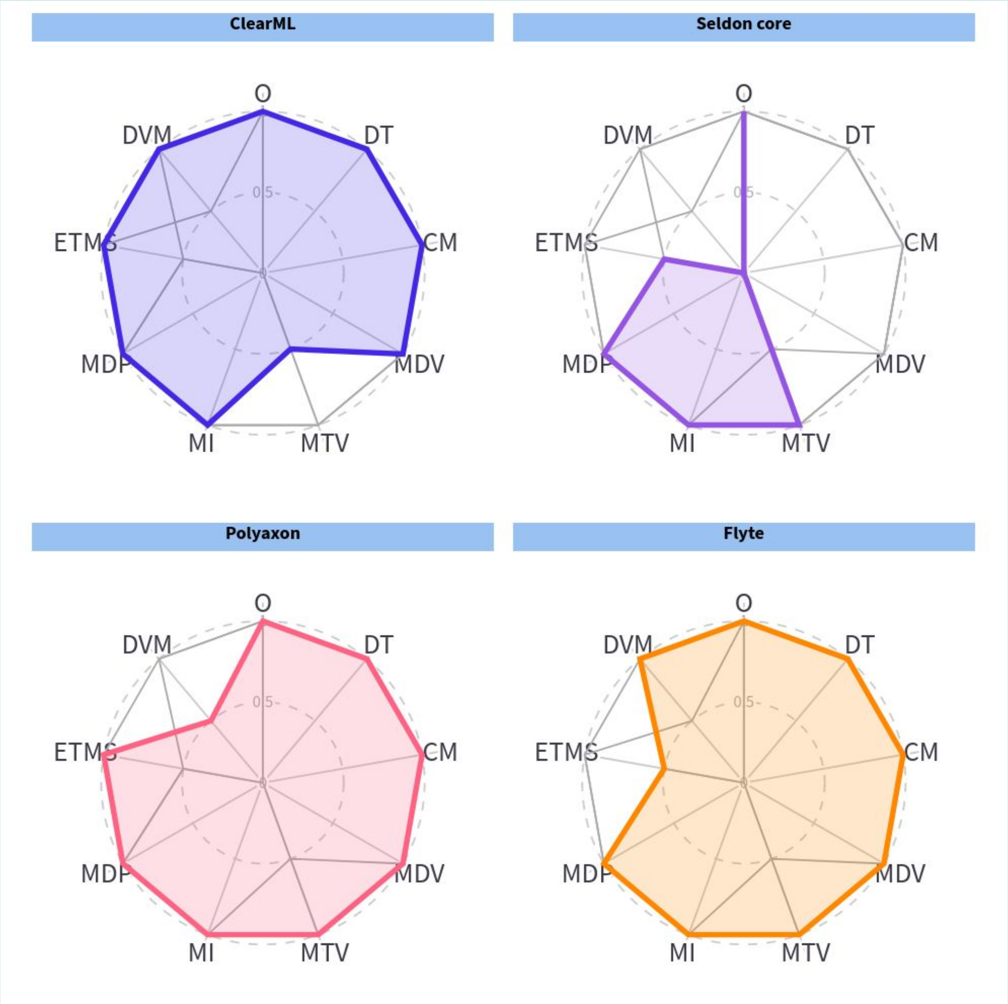
Co-funded by the European Union

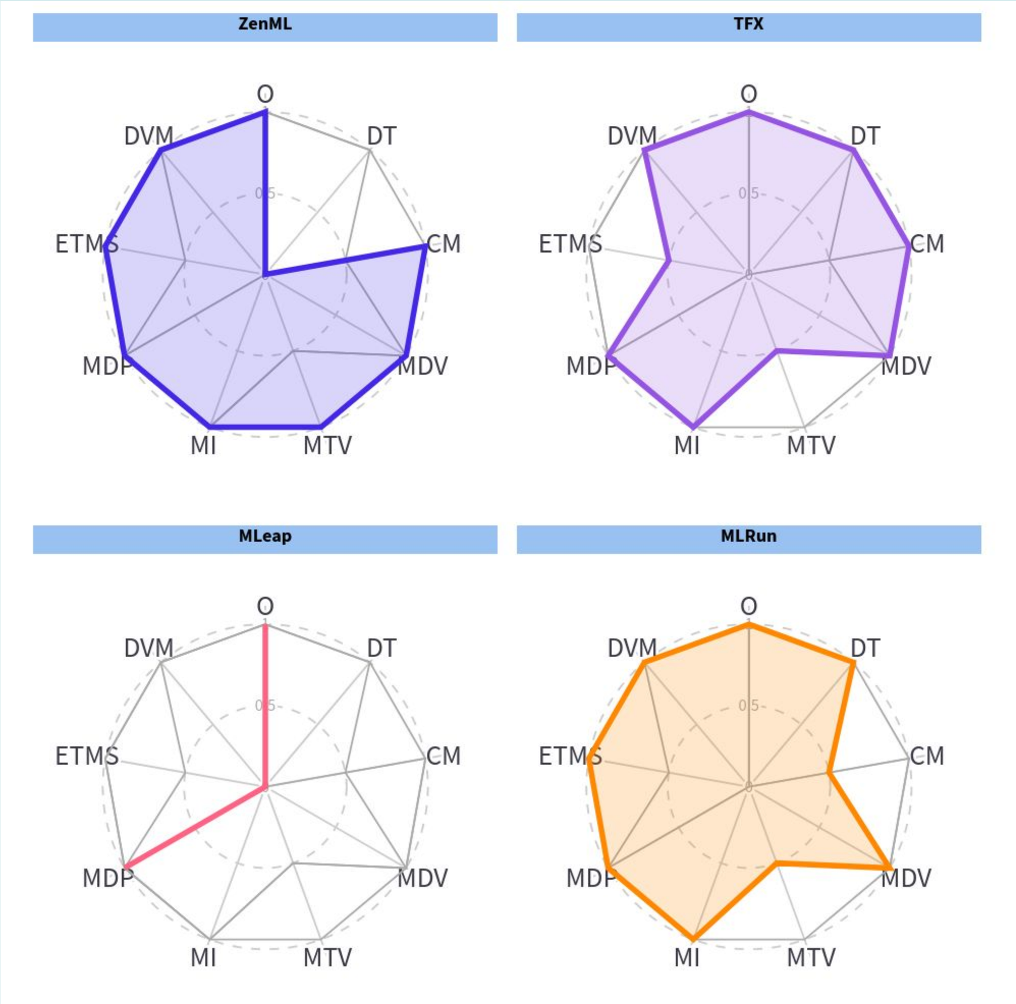


Abbreviations of the Features
O [Orchestration]
DT [Distributed Training]
CM [Code Management]
MDV [Model Development]
MTV [Model Testing/Validation]
MI [Model Inference]
MDP [Model Deployment]
ETMS [Experiment Tracking and Metadata Store]
DVM [Data Versioning and Management]

Calculated degree of the category compliance for MLOps products (* [6])

Landscaping: MLOps Frameworks

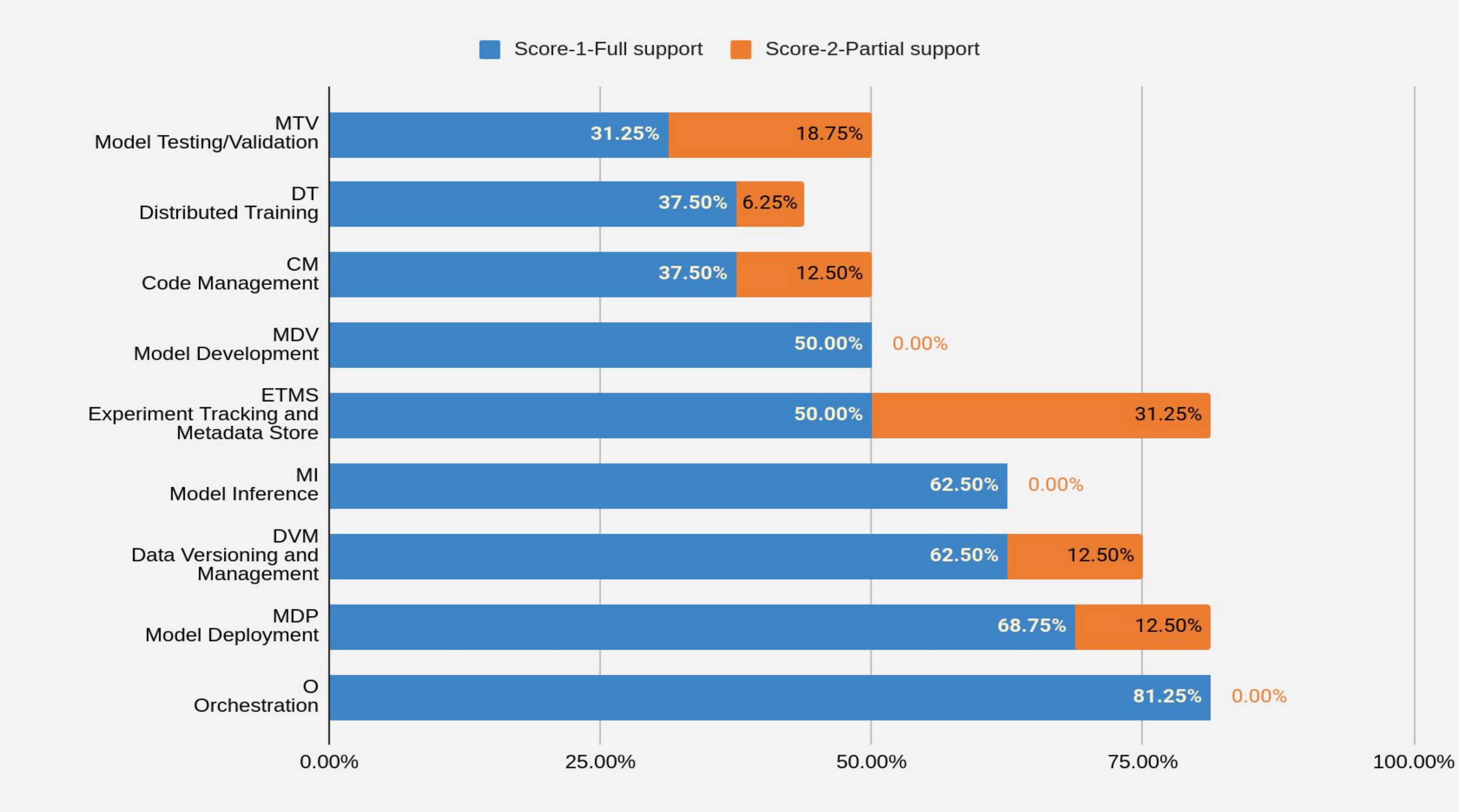




Abbreviations of the Features
O [Orchestration]
DT [Distributed Training]
CM [Code Management]
MDV [Model Development]
MTV [Model Testing/Validation]
MI [Model Inference]
MDP [Model Deployment]
ETMS [Experiment Tracking and Metadata Store]
DVM [Data Versioning and Management]

Calculated degree of the category compliance for MLOps products (* [6])

Comparison Results of Open-Source MLOps Products



Calculated percentage score of category support levels

Summary

- Products have varying levels of support for specific features
- Some Features are still in the development plan across the MLOps products.
- Select the best MLOps tool based on the requirements
- Continuous monitoring and evaluation of the evolving landscape of MLOps tools is recommended

Conclusions

- Improved Efficiency: MLOps enables automation and standardization of ML workflows, leading to faster model development, deployment, and iteration.
- Scalability: MLOps facilitates the deployment of ML models at scale, allowing organizations to handle larger datasets and serve more users.
- Reproducibility: MLOps ensures that ML experiments and results can be reproduced, providing transparency and accountability.
- Challenges: MLOps requires setting up and managing a complex infrastructure -data storage, compute resources, and orchestration frameworks.

References

1. https://nealanalytics.com/wp-content/uploads/2020/07/MLOps-Datasheet.pdf
2. https://techcommunity.microsoft.com/t5/ai-machine-learning-blog/mlops-maturity-model-with-azure-machine-learning/ba-p/3520625
3. D6.1: https://zenodo.org/record/7635453
4. D3.1: https://zenodo.org/record/7635430
5. MLFlow test instance: http://mlflow.dev.ai4eosc.eu
6. *https://public.flourish.studio/visualisation/13982942/ (interactive radar charts visualisations)
7. https://github.com/IFCA/frouros
8. AI Infrastructure Landscape, 2022 https://ai-infrastructure.org/ai-infrastructure-landscape/ . Accessed 01.03..2022.

AI4 | eosc

Co-funded by
the European Union

AI4EOSC

ai4eosc-wp6@listas.csic.es

ai4eosc.eu

Thank you! Any questions?