

Advanced AI for scientists: the AI4EOSC platform approach

Judith Sáinz-Pardo Díaz and Ignacio Heredia Cachá

Advanced Computing and e-Science group
Institute of Physics of Cantabria (IFCA) - CSIC | UC



Funded by
the European Union

11 | 06 | 2024 by J. Sáinz-Pardo and I. Heredia



AI4



Artificial Intelligence for the EOSC

- Evolution of the DEEP Hybrid DataCloud platform
- Runs September 1st 2022 – August 2025 (36 months)
- 7 academic + 2 SME + 1 non-profit organization

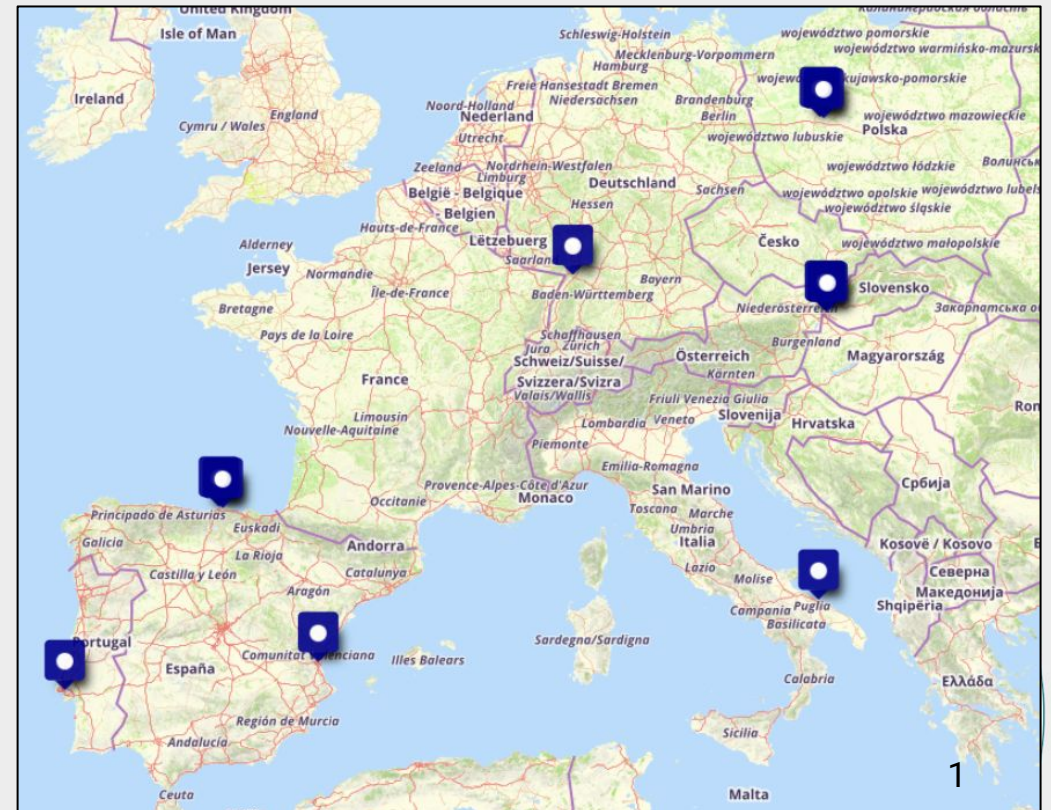
Advanced features for distributed, federated, composite learning, metadata provenance, MLOps, event-driven data processing, and provision of AI/ML/DL services

- Funding: 5M€
- 3 workshops on AI, image processing, federated learning
- 1 external users open call
- 8 peer reviewed publications in high impact journals
- 2 peer reviewed publications in high impact conferences
- Collaboration with several EU funded and INFRAEOSC projects



Funded by the European Union

11 | 06 | 2024 by J. Sáinz-Pardo and I. Heredia



AI4EOSC main objectives

1. Feature rich **services and platform** to build and deploy **custom AI applications in the EOSC**
2. Support for **building AI systems on distributed datasets**, with a particular focus on **federated learning**
3. Services to **compose AI tool workflows**, enabling the development of complex data-driven AI applications
4. **AI Exchange Hub** in the context of the EOSC, enhancing and increasing the application offer currently available
5. **Extend** the service offer and the **capabilities** being offered through the **EOSC portal**, with focus on AI

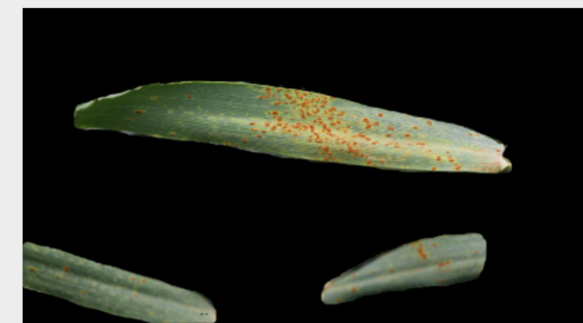
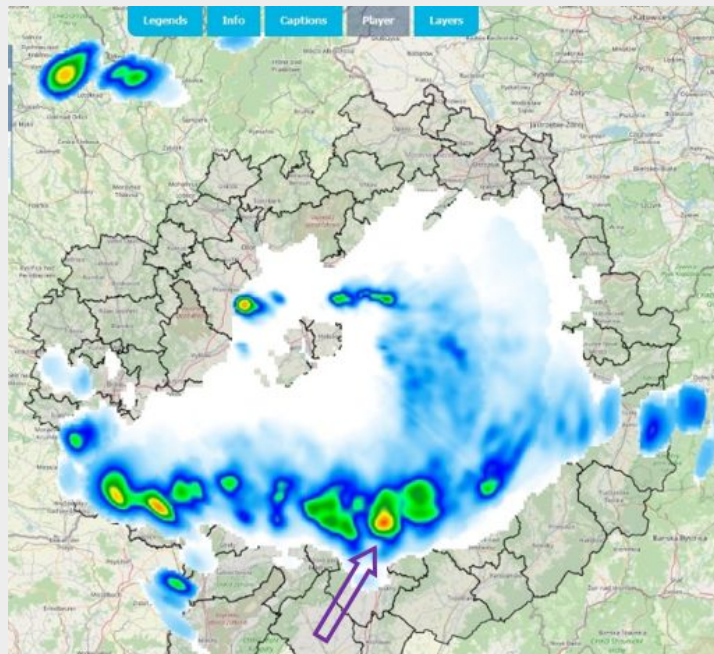


AI4EOSC use cases:

Agrometeorology

Integrated plant protection

Automated thermography



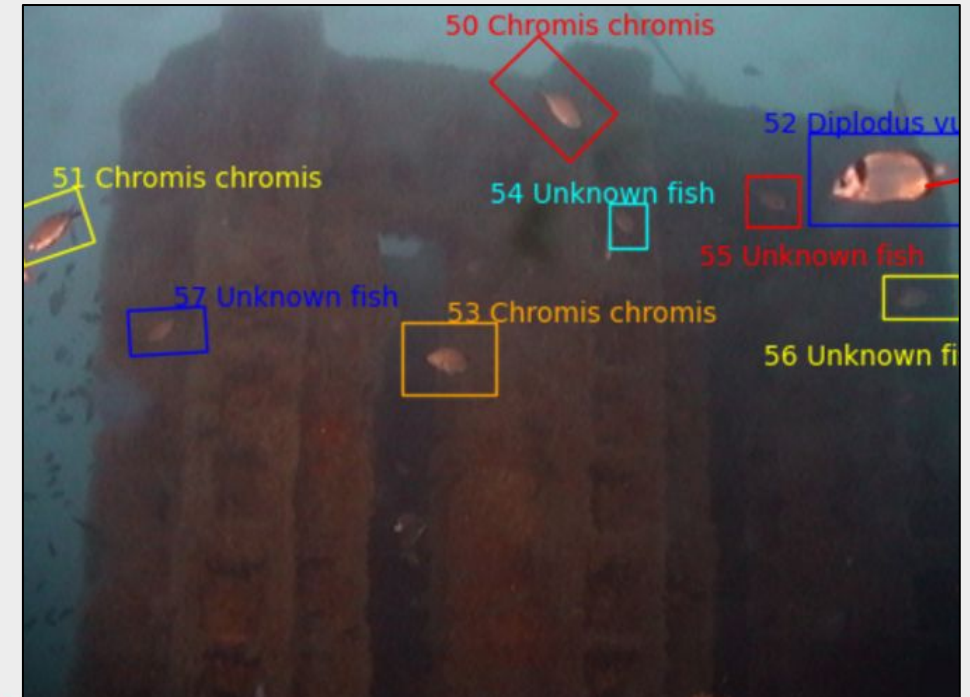
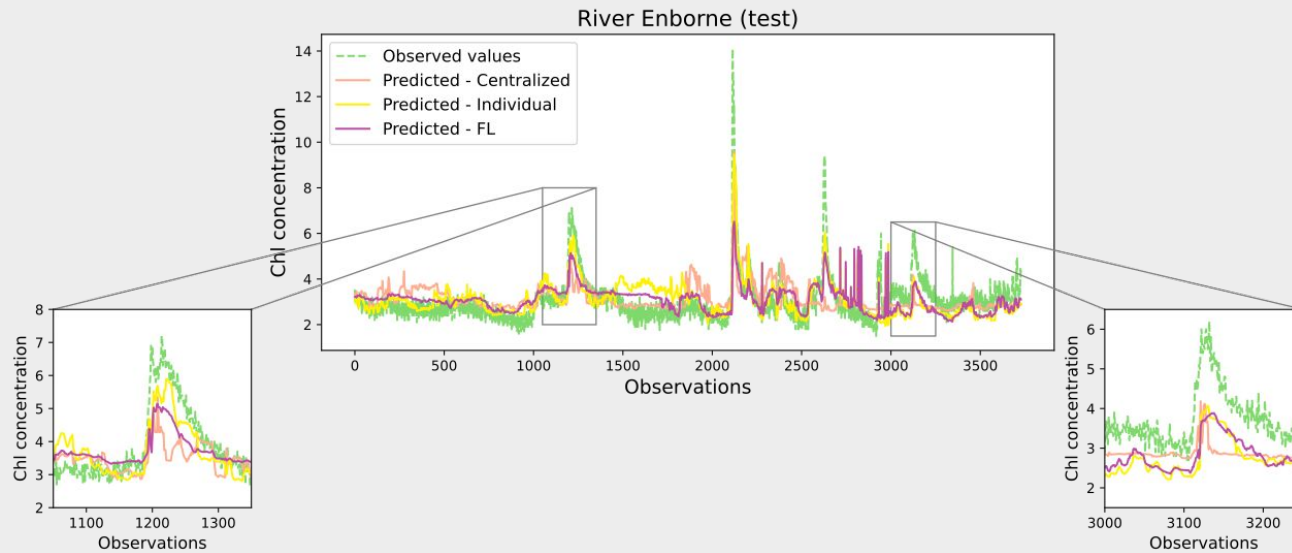
Funded by
the European Union

11 | 06 | 2024 by J. Sáinz-Pardo and I. Heredia



Water quality: predicting Chl concentration

Fish detection and classification



Sáinz-Pardo Díaz, Judith, María Castrillo, and Álvaro López García. "Deep learning based soft-sensor for continuous chlorophyll estimation on decentralized data." Water Research 246 (2023): 120726. <https://doi.org/10.1016/j.watres.2023.120726>

Martinez, Enoc and Valentin Kozlow. OBSEA Fish detector. <https://github.com/EnocMartinez/obsea-fish-detection>

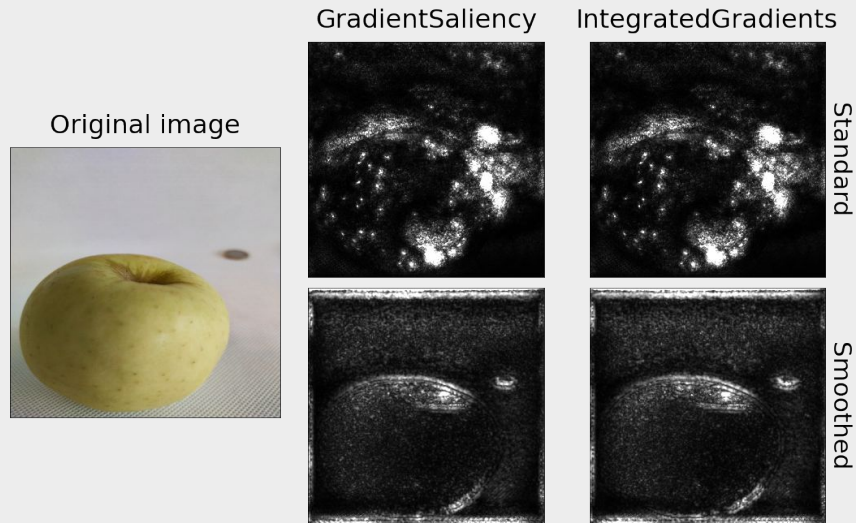


Funded by
the European Union



Medical imaging using federated learning

AI for precise weight measurement of fruits

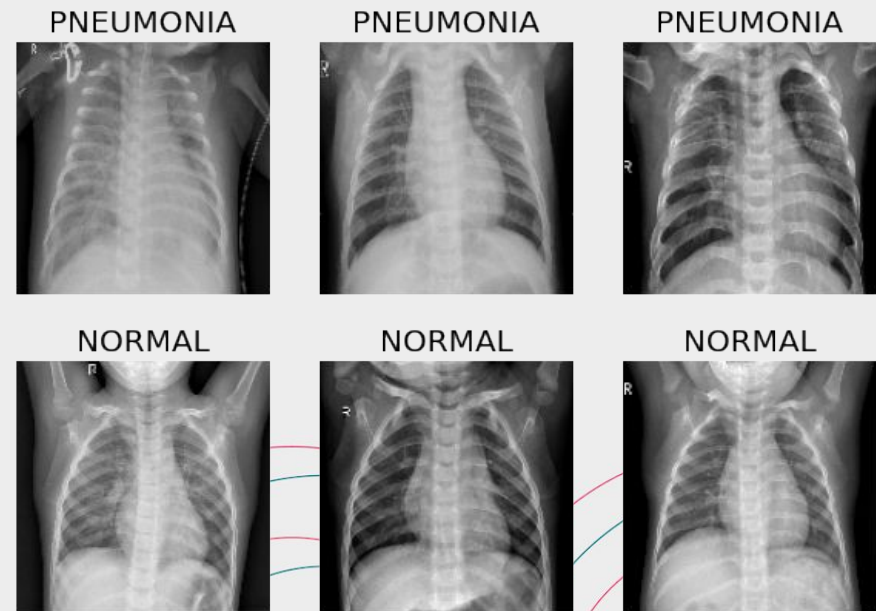


Izquierdo, Pablo, from CSIC DigitalAlimenta project.
<https://digitalalimenta.csic.es/>

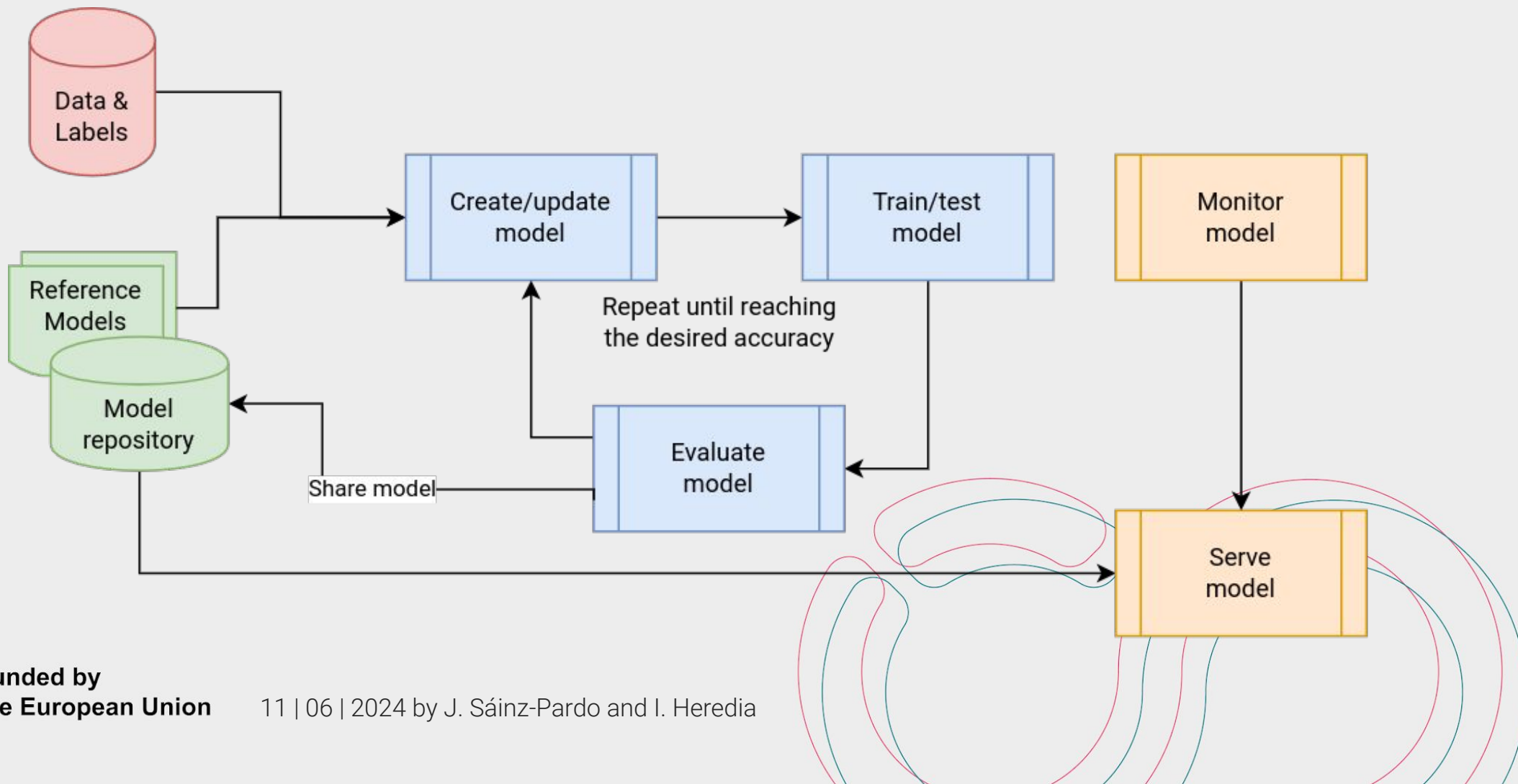


Funded by
 the European Union

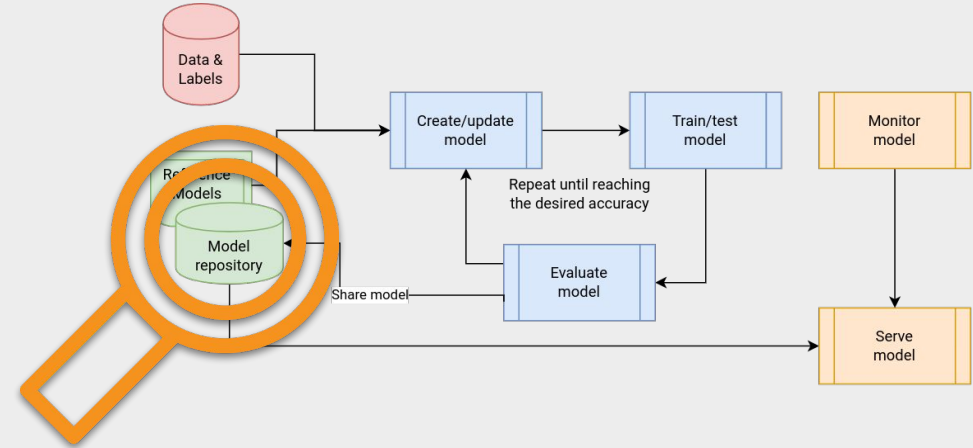
Sáinz-Pardo Díaz, Judith, and López García, Álvaro. "Study of the performance and scalability of federated learning for medical imaging with intermittent clients." *Neurocomputing* 518 (2023): 142-154.
<https://doi.org/10.1016/j.neucom.2022.11.011>




The Machine Learning Lifecycle




Model marketplace





AI4 |  eosc

- Dashboard
- Marketplace
- Deployments

Useful links 


- Identity and Access
- AI4EOSC documentation
- Project page
- Storage
- Status
- Experiment tracking

 AI4 |  eosc

The AI4EOSC dashboard is a service provided by CSIC, co-funded by [AI4EOSC](#)


[Terms of use](#) [Privacy policy](#)

v2.1.0

Marketplace 


Modules Tools

Judith Sáinz-Pardo Díaz

 AI4OS Development Environment


This is a Docker image for developing new modules

New deployment


 Dogs breed detector

Identify a dogs breed on the image (133 known breeds)


Trainable Inference Pre-trained

 DEEP OC Massive Online Data Streams

Deep learning for proactive network monitoring and security protection.


 DEEP OC Retinopathy Test

A Tensorflow model to classify Retinopathy.

 Train an image classifier


Train your own image classifier with your custom dataset. It comes also pretrained on the 1K ImageNet classes.


Trainable Inference Pre-trained

 Plants species classifier

Classify plant images among 10K species from the iNaturalist dataset.

Trainable Inference Pre-trained

 Upscale multispectral satellites

 Speech keywords classifier

Plants species classifier

Classify plant images among 10K species from the iNaturalist dataset.

Build status: build passing | License: Apache 2.0 | Created: 2019-01-01

The deep learning revolution has brought significant advances in a number of fields [1], primarily linked to image and speech recognition. The standardization of image classification tasks like the [ImageNet Large Scale Visual Recognition Challenge](#) [2] has resulted in a reliable way to compare top performing architectures.

The use of deep learning for plant classification is not novel [3, 4] but has mainly focused in leaves and has been restricted to a limited amount of species, therefore making it of limited use for large-scale biodiversity monitoring purposes.

This Docker container contains a trained Convolutional Neural network optimized for plant identification using images. The architecture used is an Xception [5] network using Keras on top of Tensorflow. A detailed article about this network and the results obtained with it can be found in [6].

The PREDICT method expects an RGB image as input (or the url of an RGB image) and will return a JSON with the top 5 predictions.

The original training dataset was the great collection of images which are available in [PlantNet](#) under a Creative-Common AttributionShareAlike 2.0 license. It consists of around 250K images belonging to more than 6K plant species of Western Europe. These species are distributed in 1500 genera and 200 families.

A new iteration of the application has been trained using plant images from [iNaturalist](#). This dataset has around 4.4M observations with 7M images from 58K worldwide species. We have restricted our training to the 10K most popular species.



Categories

- tensorflow
- docker
- deep learning
- trainable
- inference
- pre-trained
- image classification
- api-v2

Additional Resources

Get the code

-  Github
-  Dockerhub

Get the data

-  Dataset

Found an issue?

-  Report issue

Deploy via the IM

Train module



Funded by
the European Union

11 | 06 | 2024

Create your own model

Configure training: AI4OS Development Environment

Marketplace / Plants species classifier / Train Show help

1 General configuration — 2 **Hardware configuration** — 3 Storage configuration

Hardware options

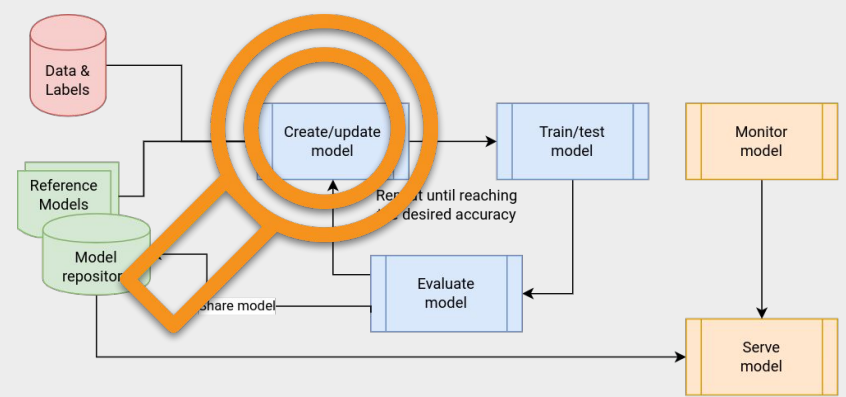
Number of CPUs: 8

Number of GPUs: 1

GPU model: Tesla V100-PCIE-32GB

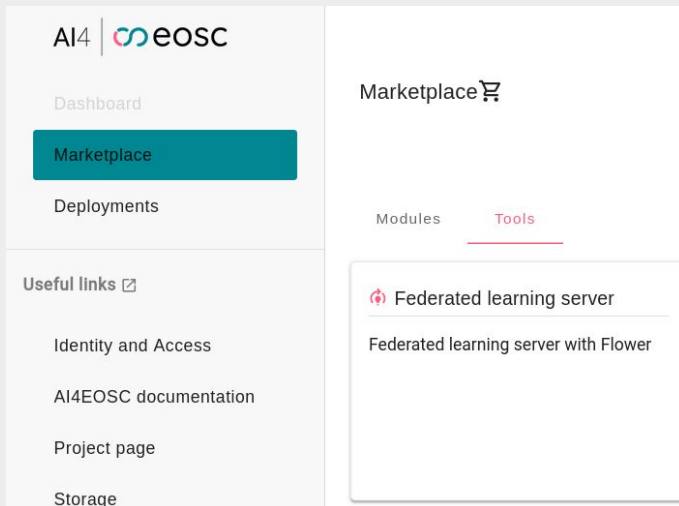
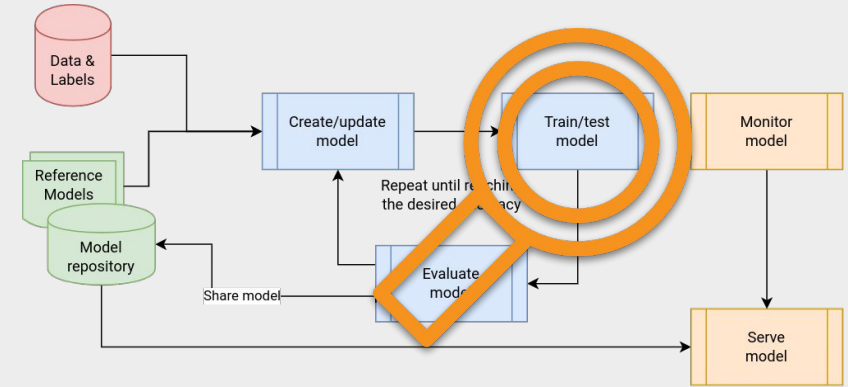
RAM memory (in MB): 16000

Disk memory (in MB): 10000



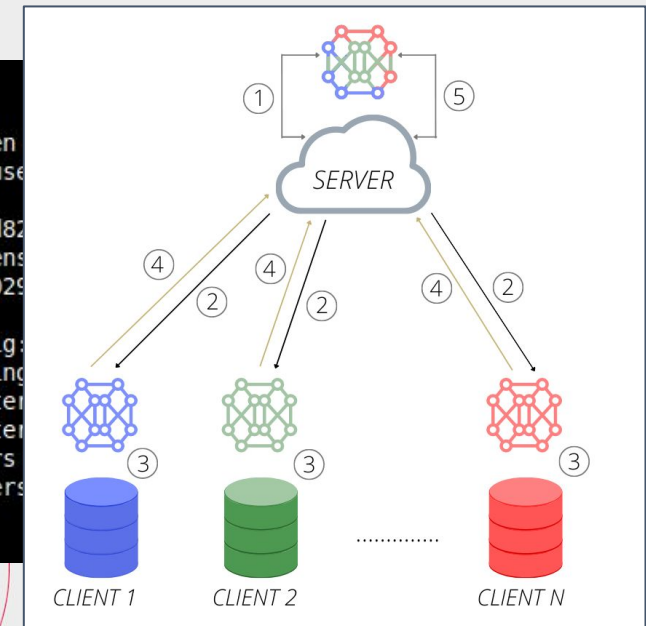

Training: federated learning

- Collaborative and decentralized approach to build ML models
 - No need to centralize a dataset (i.e. technical or privacy restrictions)
- Management of experiments through platform dashboard
- Participating clients both within AI4EOSC platform or external (with authentication)



```

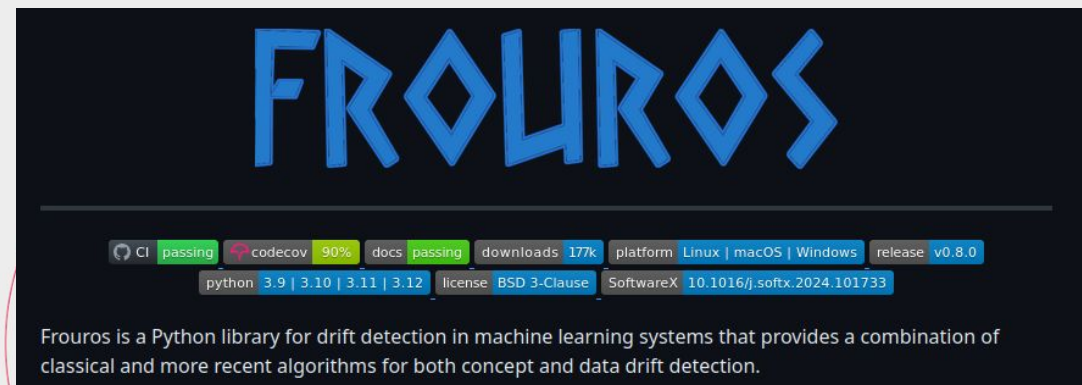
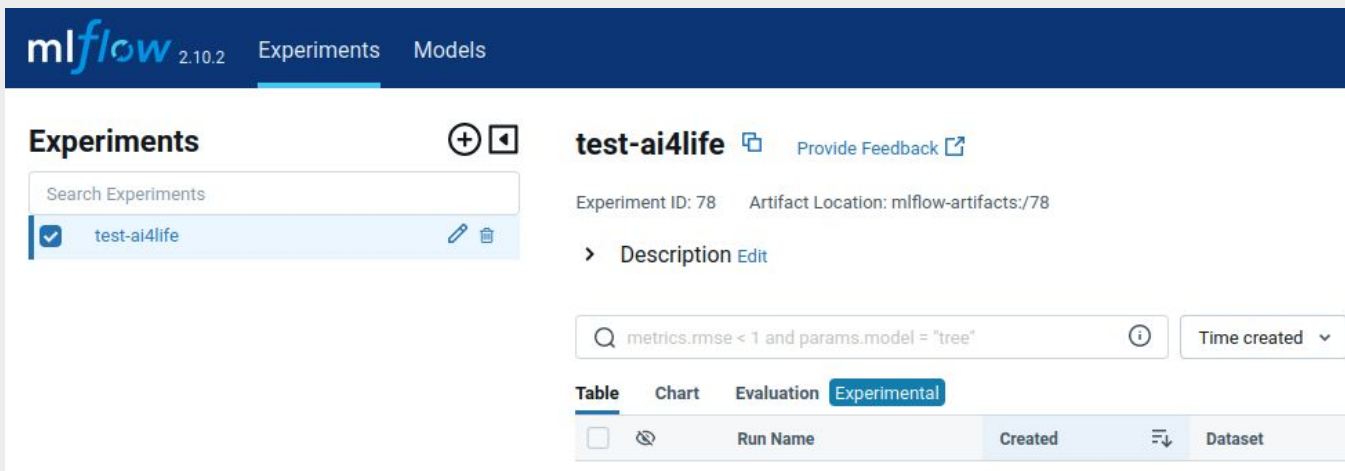
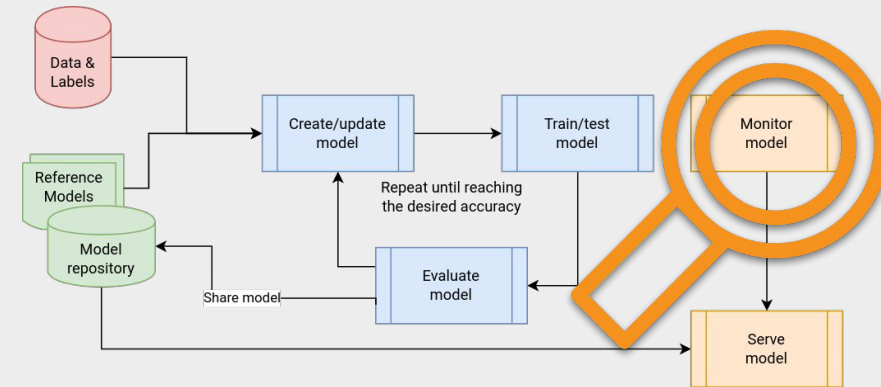
root@8878206726fb:/srv# cd federated-server/fedserver
root@8878206726fb:/srv/federated-server/fedserver# python3 server.py
INFO flwr 2024-04-18 14:10:23,381 | vault.py:76 | Configured Vault Bearer token
INFO flwr 2024-04-18 14:10:23,381 | vault.py:77 | Reading tokens stored in: 'use
a3f-0242ac130005/federated'
Getting tokens from Vault -> users/7d7a87545b700b38b54e2b5b4713084fd2b8d7e5ed82
INFO flwr 2024-04-18 14:10:23,790 | vault.py:79 | Configured Vault Bearer tokens
86bf6d2843ad31a8879a5fab0c1318', '6707ded3dab865271f4a5ac637601cbb38269b808f2029
INFO flwr 2024-04-18 14:10:23,790 | server.py:80 | Token interceptor created
INFO flwr 2024-04-18 14:10:23,791 | app.py:158 | Starting Flower server, config:
INFO flwr 2024-04-18 14:10:23,800 | app.py:172 | Flower ECE: gRPC server running
INFO flwr 2024-04-18 14:10:23,800 | server.py:91 | Initializing global parameter
INFO flwr 2024-04-18 14:10:23,800 | server.py:282 | Requesting initial parameter
INFO flwr 2024-04-18 14:10:49,423 | server.py:288 | Received initial parameters
INFO flwr 2024-04-18 14:10:49,424 | server.py:93 | Evaluating initial parameters
INFO flwr 2024-04-18 14:10:49,424 | server.py:106 | FL starting
    
```



Funded by
the European Union

Monitor the model: MLOps and drift detection

- Monitoring of models in production is not enough
 - Model learns from data, data is not stationary
 - Concept learnt by the model may change over time
- Data and concept drift detection → essential to build more robust models
- Frouros: state-of-the-art Python library for drift detection in ML problems: <https://github.com/IFCA/frouros>
- MLOps is an engineering practice that aims to automate and streamline the ML lifecycle
- MLflow tracking server: <https://mlflow.cloud.ai4eosc.eu>



Serve the model: inference

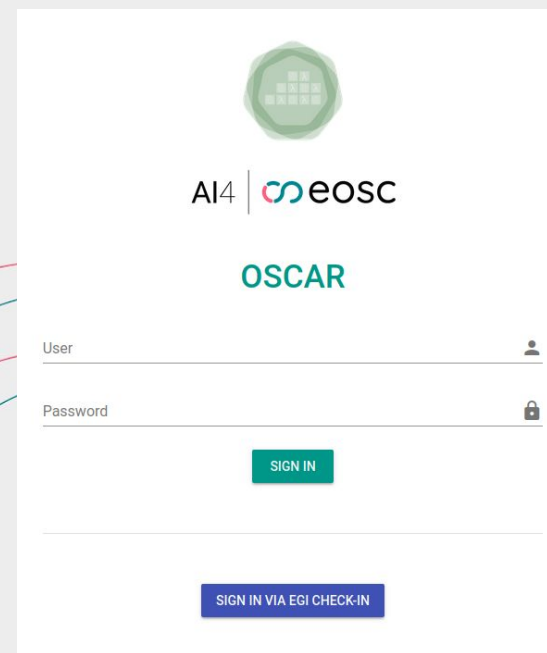
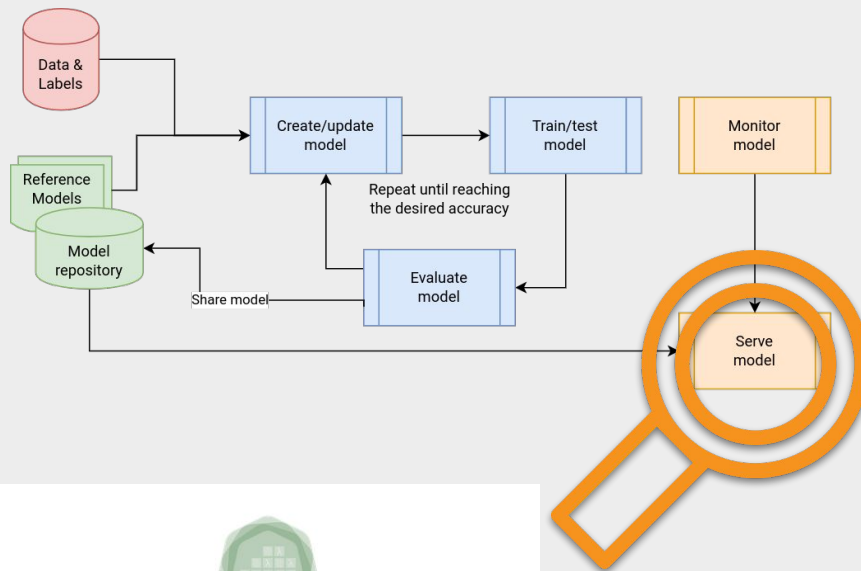
<https://inference.cloud.ai4eosc.eu/ui/>

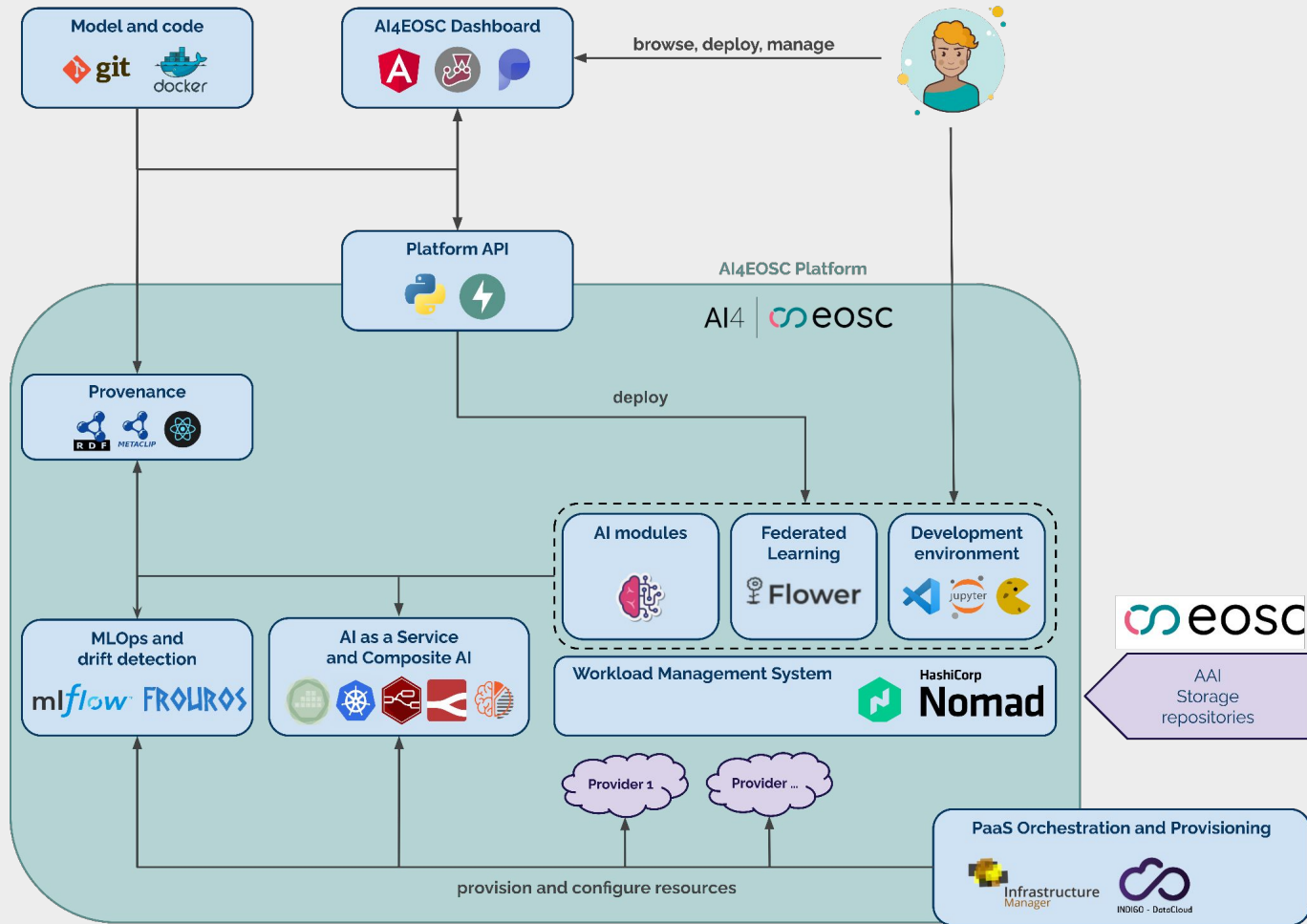
We use [OSCAR](#) to serve AI models for inference (AI as a Service).

It supports two serverless event-driven execution modes:

- **Asynchronous mode:** Files uploaded to the object-store automatically trigger the invocation of a data-processing script, that is run inside a container (out of user-defined Docker image) within a scalable Kubernetes cluster (e.g. batch jobs).
- **Synchronous mode:** Scalable HTTP-based endpoints (based on KNative). Direct requests to the model.

On top, we support building AI workflows using [Flowfuse](#) and [Elyra](#) through [AI4Compose](#).



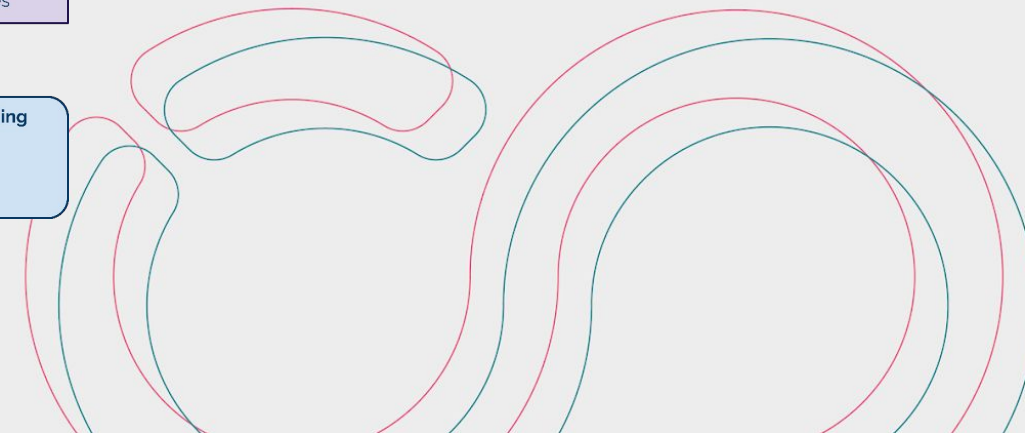


Interactive C4 diagrams available [here](#).



Funded by the European Union

11 | 06 | 2024 by J. Sáinz-Pardo and I. Heredia



AI4



Showcasing the AI4EOSC platform

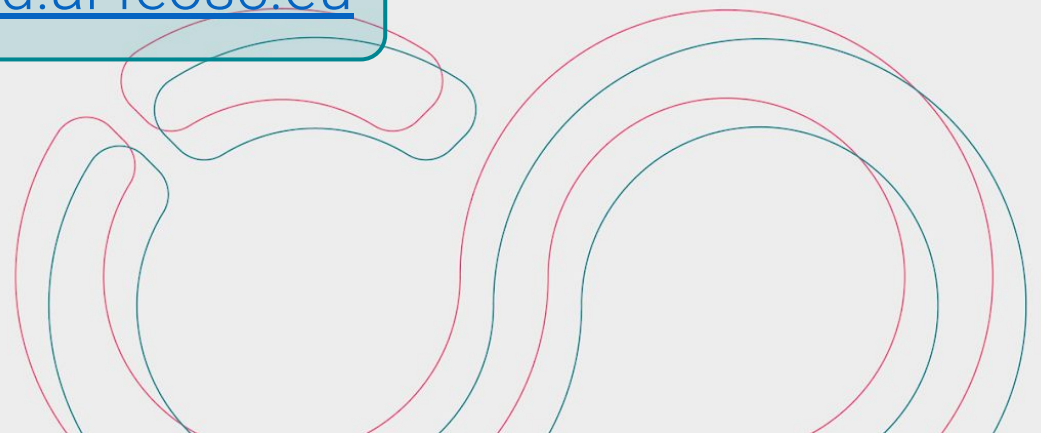


<https://dashboard.cloud.ai4eosc.eu>



Funded by
the European Union

11 | 06 | 2024 by J. Sáinz-Pardo and I. Heredia



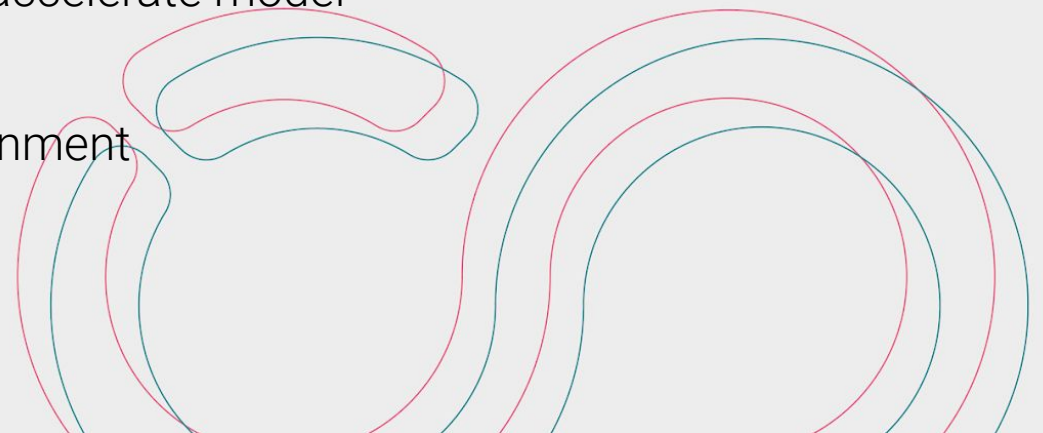
AI4EOSC empowers scientific research by...

- Providing users with advanced AI tools:
 - Model retraining (iterative learning, fine tuning)
 - Federated learning (including client authentication)
 - Parallel training in multiple GPUs (distributed training - data parallelism)
 - Model monitoring: MLOps, drift detection
 - Model inference
- Providing a simple and intuitive IDE for developing AI/ML/DL models (VSCode or JupyterLab)
- Allowing seamless access to computational resources to accelerate model development
- Deploying your models in production in a serverless environment

More info:

<https://ai4eosc.eu/>

<https://docs.ai4os.eu/>



Thank you for your attention!

Judith Sáinz-Pardo Díaz
Ignacio Heredia Cachá

sainzpardo@ifca.unican.es
iheredia@ifca.unican.es

Álvaro López García
(project coordinator)

aloga@ifca.unican.es

