iMagine: Best practices for suppliers of image collections and analysis tools in aquatic sciences

E. Azmi[1], K. Alibabaei[1], V. Kozlov[1], A. Lopez Garcia[2], D. Schaap[3], G. Sipos[4]

[1] KIT, [2] CSIC, [3] MARIS, [4] EGI

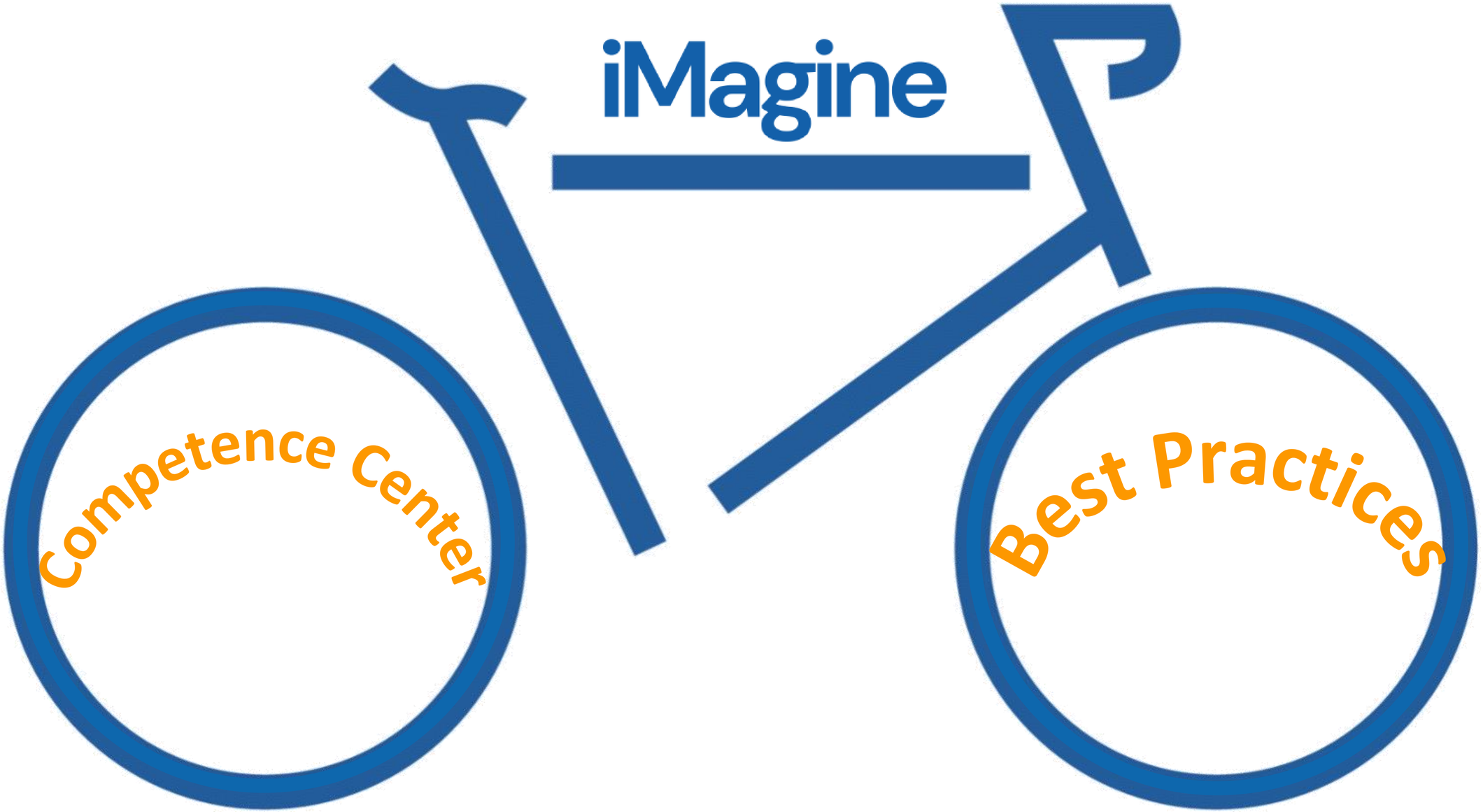
EGI2024

September 30 - October 4

iMagine receives funding from the European Union's Horizon Europe research and innovation program under grant agreement No. 101058625

Funded by the European Union

iMagine



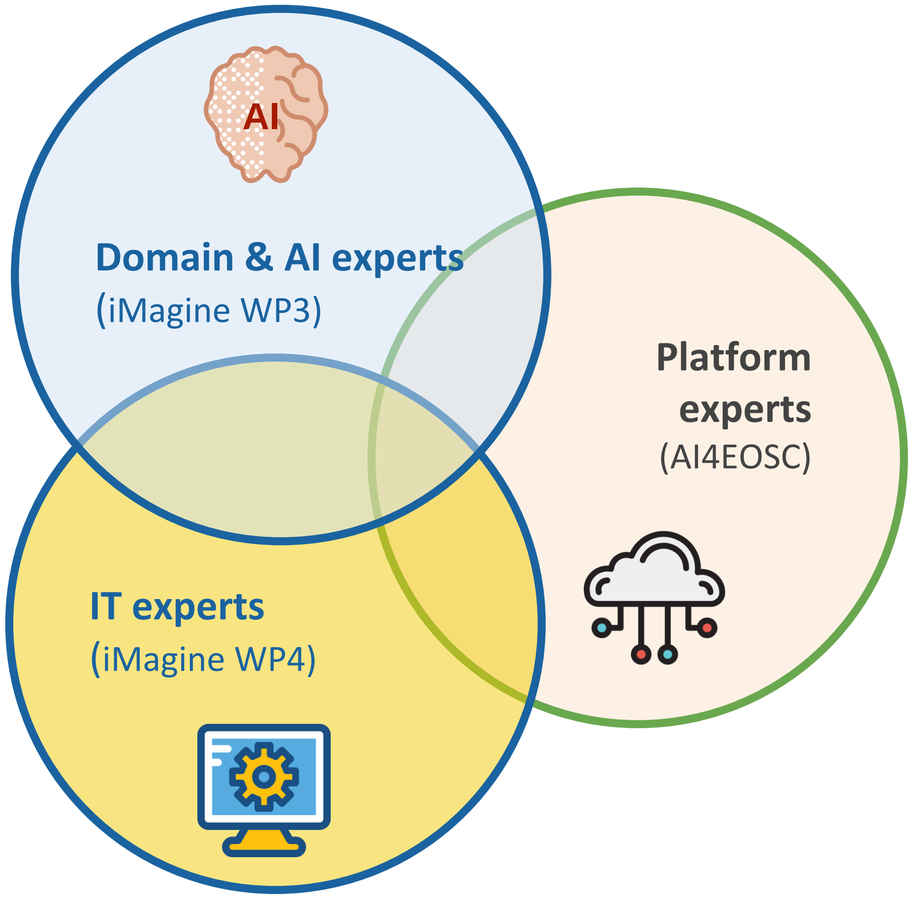
Competence Center:

Introduction

- AI, IT and domain experts from the iMagine:
  - WP3: marine and freshwater application experts
  - WP4: AI framework / IT experts
- Monitoring and supporting of use cases through:
  - Regular meetings
  - Annual competence workshops
- Provide input for a “Best Practices” document
- Support in standardization and improving quality of datasets

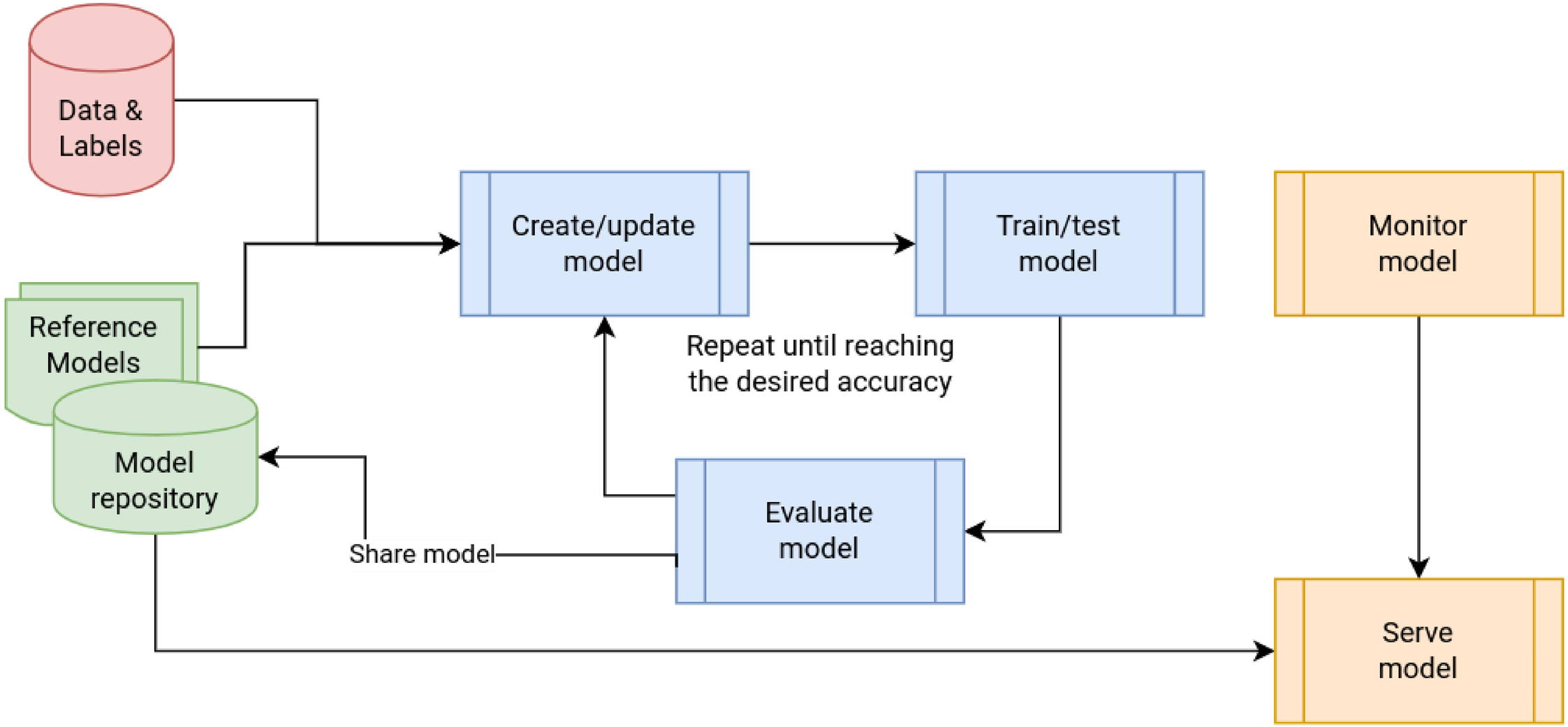
Competence Center:

Synergies



Competence Center:

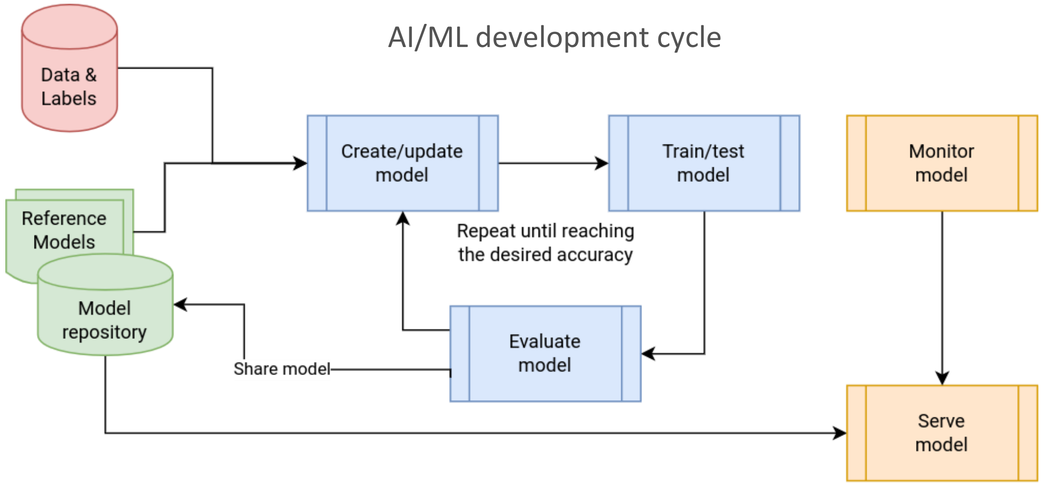
Support Whole AI/ML Development Cycle



Competence Center:

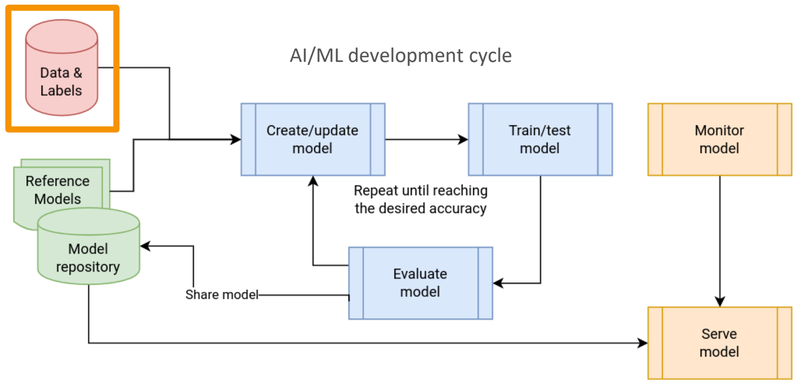
Achieved Best Practices

- Deep Learning
- Annotation Tools
- Data Repositories and Open-source Dataset for Marine Applications
- Preprocessing Techniques
- Performance Metrics and Evaluation Methods
- Tools for Monitoring Model Performance
- Data Biases and Fairness in Aquatic Science Models and Data
- Model Delivery
- AI Model Drift Tools
- Use Case Experiences



Best Practices:

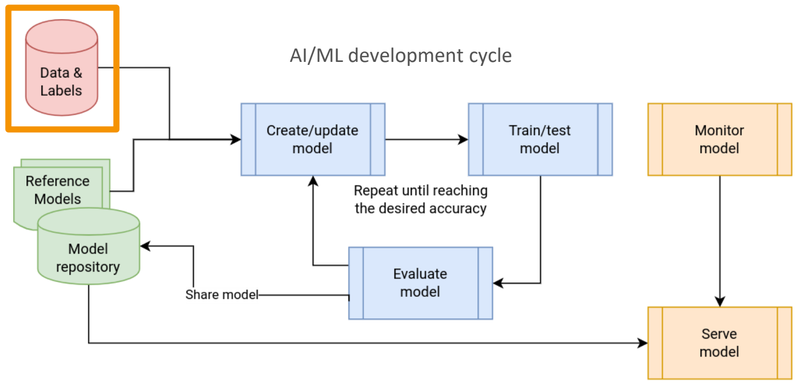
Data and Labels



- Annotation Tools
  - CVAT: user-friendly interface, robust annotation features, and flexibility
  - Label Studio: more complex to use and install locally
  - BIIGLE: web-based for marine image annotation
  - RoboFlow: cloud-based platform with data augmentation
  - LabelBox: customizable for different data types

Best Practices:

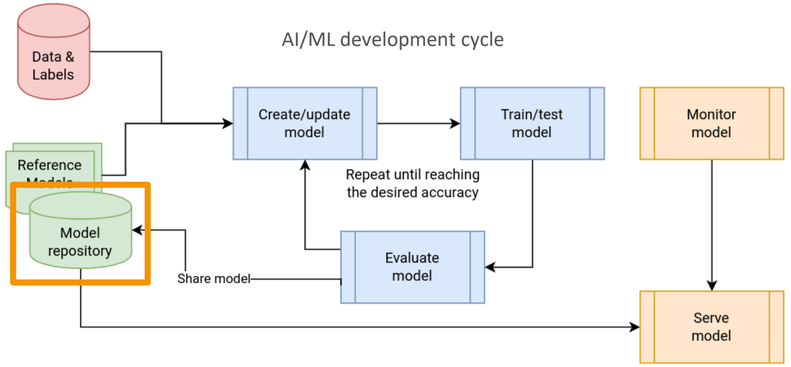
Data and Labels



- Preprocessing Techniques
  - Data Cleaning: removing errors and inconsistencies
  - Normalization & Standardization: scaling pixel values to same scale
  - Handling Imbalanced Data: oversampling/undersampling and class weight adjustment
  - Data Augmentation: transformations such as flipping and rotating

Best Practices:

Data Repository



- Data Repositories and Open-source Dataset for Marine Applications
  - Zenodo: robust solution, open-access platform to share, store, and manage datasets

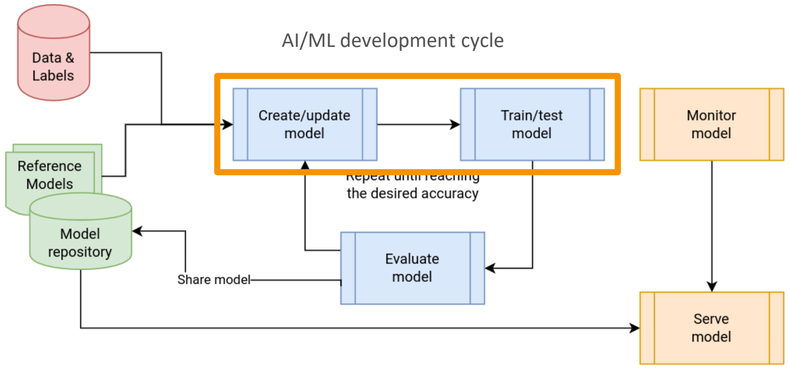
zenodo.org

- Data Biases and Fairness in Aquatic Science Models and Data
  - FAIR EVA tool: improve the FAIRness of published training datasets concerning the metadata used in Zenodo

FAIReva

Best Practices:

Create/Update/Train/Test Model

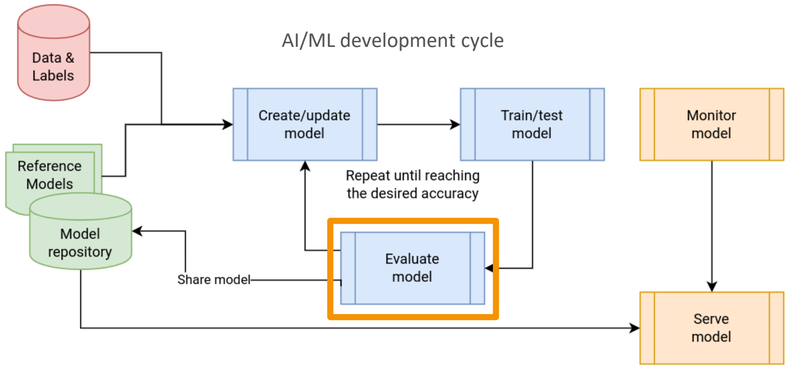


- Deep Learning Models
  - Classification
    - Mobilenet
  - Object Detection
    - YOLOv8: most commonly used for object detection, showed strong performance
  - Segmentation
    - YOLOv8
    - Mask2Former: achieved better performance in panoptic segmentation than instance segmentation with Mask R-CNN
- Training and Testing on iMagine Marketplace

Marketplace

Best Practices:

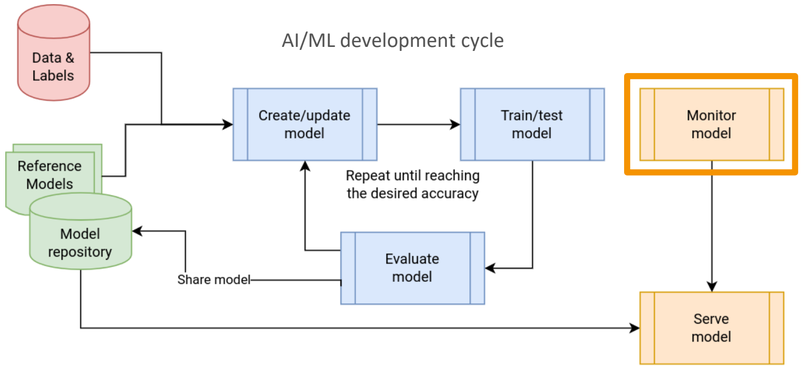
Evaluate Model



- Performance Metrics and Evaluation Methods
  - Accuracy: correct predictions over total predictions
  - Precision: the ratio of correctly predicted positive observations to the total predicted positives
  - Recall: the ratio of correctly predicted positive observations to all observations in the actual class
  - F1 score: the harmonic mean of precision and recall, useful for imbalanced dataset
  - Intersection over Union (IoU): the overlap between two boxes, with greater overlap indicating a higher IoU

Best Practices:

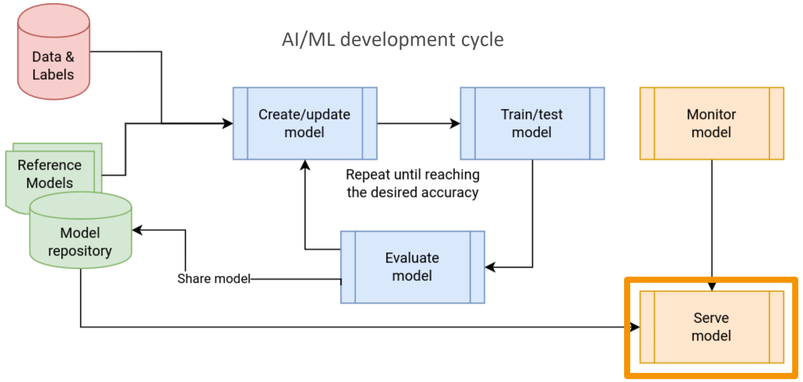
Monitoring Model Performance



- Experiment Tracking Tools
  - Mlflow: efficient management and tracking of machine learning experiments
  - Tensorboard: less functionalites, single user and self-hosted
  - Weights and Biases: self-hosted server is by default is not appropriate for production environment

Best Practices:

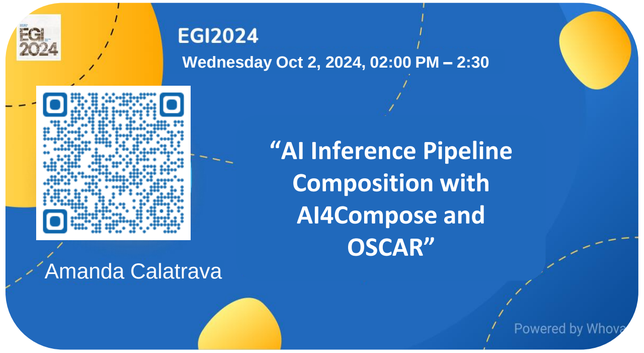
Serve Model



- Sharing and Serving AI Models via iMagine Marketplace
  - AI models are published as Docker images on the iMagine marketplace
  - Anyone authenticated can “Try” an inference endpoint (for 10 min)

Marketplace

- Model Deployment in Production
  - OSCAR: efficient and scalable deployment for running the AI model inferences



iMagine

Published October 1, 2024 | Version Under EC Review

Project deliverable | Open

iMagine D3.4 Best practices for producers and providers of image sets and image analysis applications in aquatic sciences

Alibabaei, Khadijeh[1]; Azmi, Elnaz[1]; Kozlov, Valentin[1]

Show affiliations

iMagine is a European project to serve aquatic researchers with a suite of high-performance image analysis tools equipped with artificial intelligence. To effectively achieve the objectives of the project, eight use cases in different areas of aquatic science are collaborating with the providers of the iMagine AI platform. This collaboration has yielded valuable insights and practical knowledge.

In this deliverable, we delve into the details of the best AI-based solutions for image processing in aquatic science, drawing on the extensive experience and knowledge we have gained over the course of the iMagine project. We thoroughly review the methods and tools used in the initial phase of data labelling, in the subsequent phases of model training and in the final deployment of the model as a service.

Marketplace | Image processing Tips | twitter.com | zenodo.org | imagine-ai.eu

External resources

Indexed in

OpenAIRE

Communities

EU Open Research Repository (Pilot)

iMagine project

Keywords and subjects

Best practices | AI | Image sets | Analysis

Details

DOI

DOI 10.5281/zenodo.13864197

Resource type

Project deliverable

Publisher

Zenodo

Languages

English

Rights

Creative Commons Attribution 4.0 International

Citation

Alibabaei, K., Azmi, E., & Kozlov, V. (2024). iMagine D3.4 Best practices for producers and providers of image sets and image analysis applications in aquatic sciences (Under EC Review). Zenodo. https://doi.org/10.5281/zenodo.13864197

Style APA

Thank you for your attention

imagine-ai.eu

iMagine receives funding from the European Union's Horizon Europe research and innovation program under grant agreement No. 101058625

Funded by the European Union