

Open source software MLOps platform (OSS MLOps)

Quality, scale and trust for AI
in all environments
with free, integrated and extensible toolchain

26.9.2024

Jukka Remes / Haaga-Helia UAS

Harry Souris / Silo AI

Niila Siilasjoki / University of Helsinki

SILO_{AI}

IML4E
Industrial Machine Learning for
Enterprises



 **Fraunhofer**

 **Haaga-Helia**
University of Applied Sciences



Why MLOps?

Secure transparency and meet compliance requirements

- Providing solutions and frameworks that meet regulatory compliance and audit requirements.
- Bringing business stakeholder closer to the ML products.



Why MLOps?

Secure transparency and meet compliance requirements

Trusted and reproducible AI in production - **secure AI investments**

Providing solutions and frameworks that **meet regulatory compliance and audit requirements.**

Bringing business stakeholder closer to the ML products. Enabling stakeholders to act and to set expectations on the execution of the models.

Ensure customers are always facing updated and quality AI/ML services.

- Not risking enterprise's reputation by operating outdated models in production environments.
- AI/ML services adhering to business and data changes. A result of the frequent and automated retraining and deployment of models.
- Quality assurance through automating the process of testing and validating and monitoring models and data.



Why MLOps?

Secure transparency and meet compliance requirements

Trusted and reproducible AI in production - **secure AI investments**

Providing solutions and frameworks that **meet regulatory compliance and audit requirements.**

Bringing business stakeholder closer to the ML products. Enabling stakeholders to act and to set expectations on the execution of the models.

Ensuring customers are always facing updated and quality AI/ML services.

Not risking enterprise's reputation by operating outdated models in production environments.

AI/ML services adhering to business and data changes. A result of the **frequent and automated** retraining and deployment of models.

Quality assurance through automating the process of testing and validating and monitoring models and data.

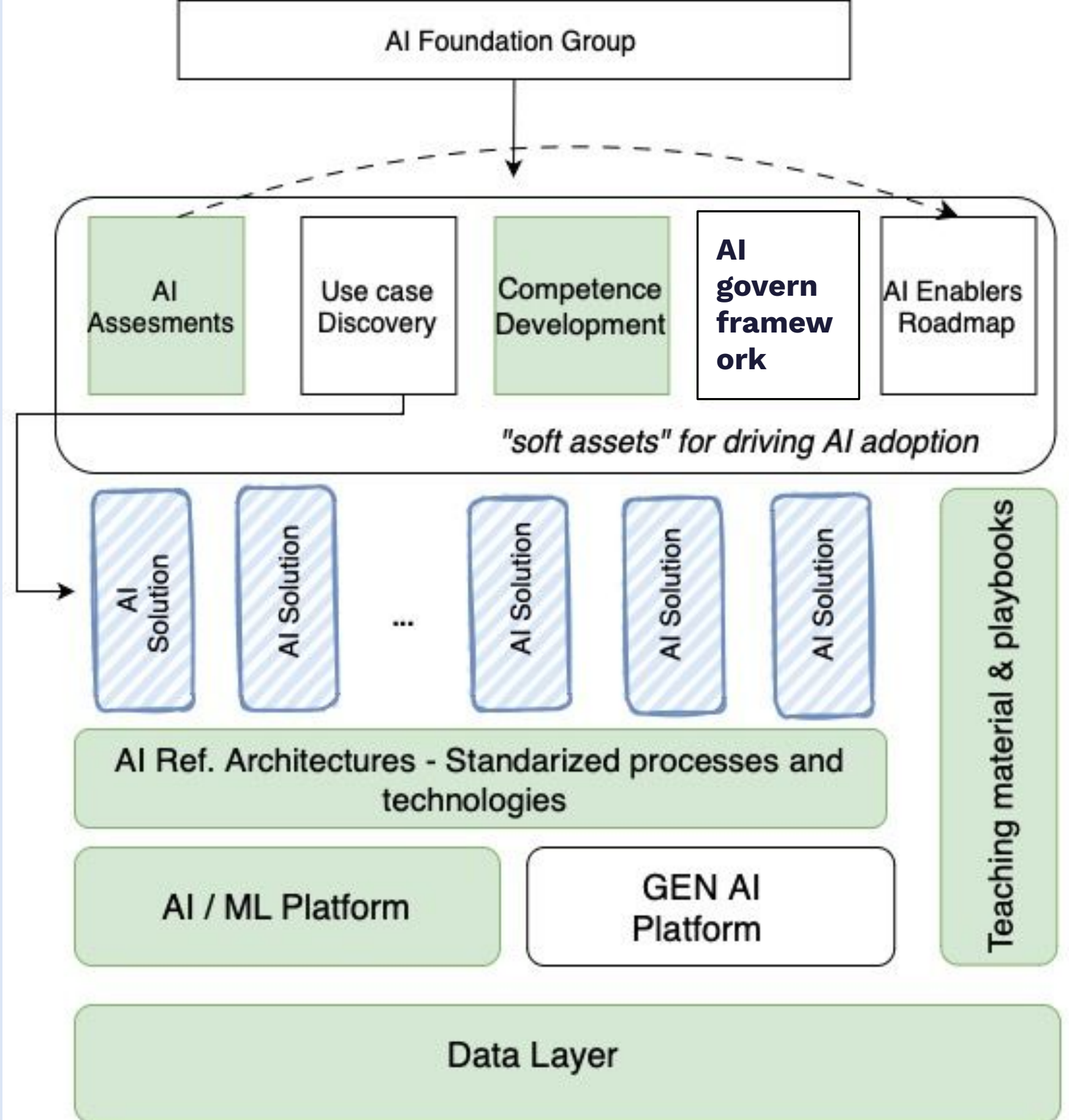
Increase productivity - enabling scalable and sustainable AI/ML utilization.

- Reducing operation and maintenance costs (>30% reduced costs). Result of utilising MLOps and Data platforms across organisations
- Maximise return on AI investments. From ideation to production with fast and controlled steps.
- Streamline AI work effort across organisation.



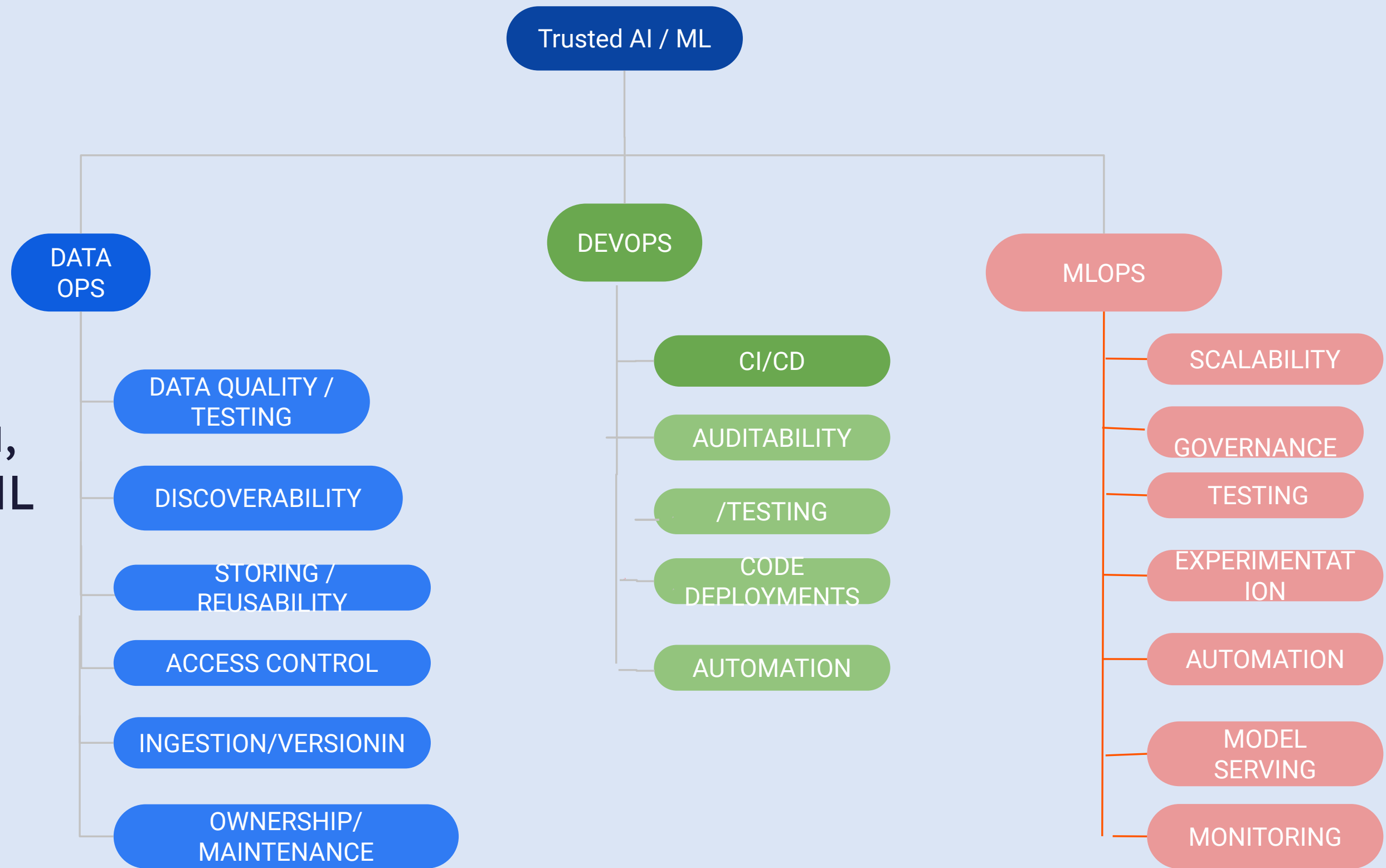
How MLOps?

A holistic approach is needed for capturing business technology and compliance requirements.



How MLOps ? Looking at the technology side...

A platform is needed that addresses data, software and ML challenges

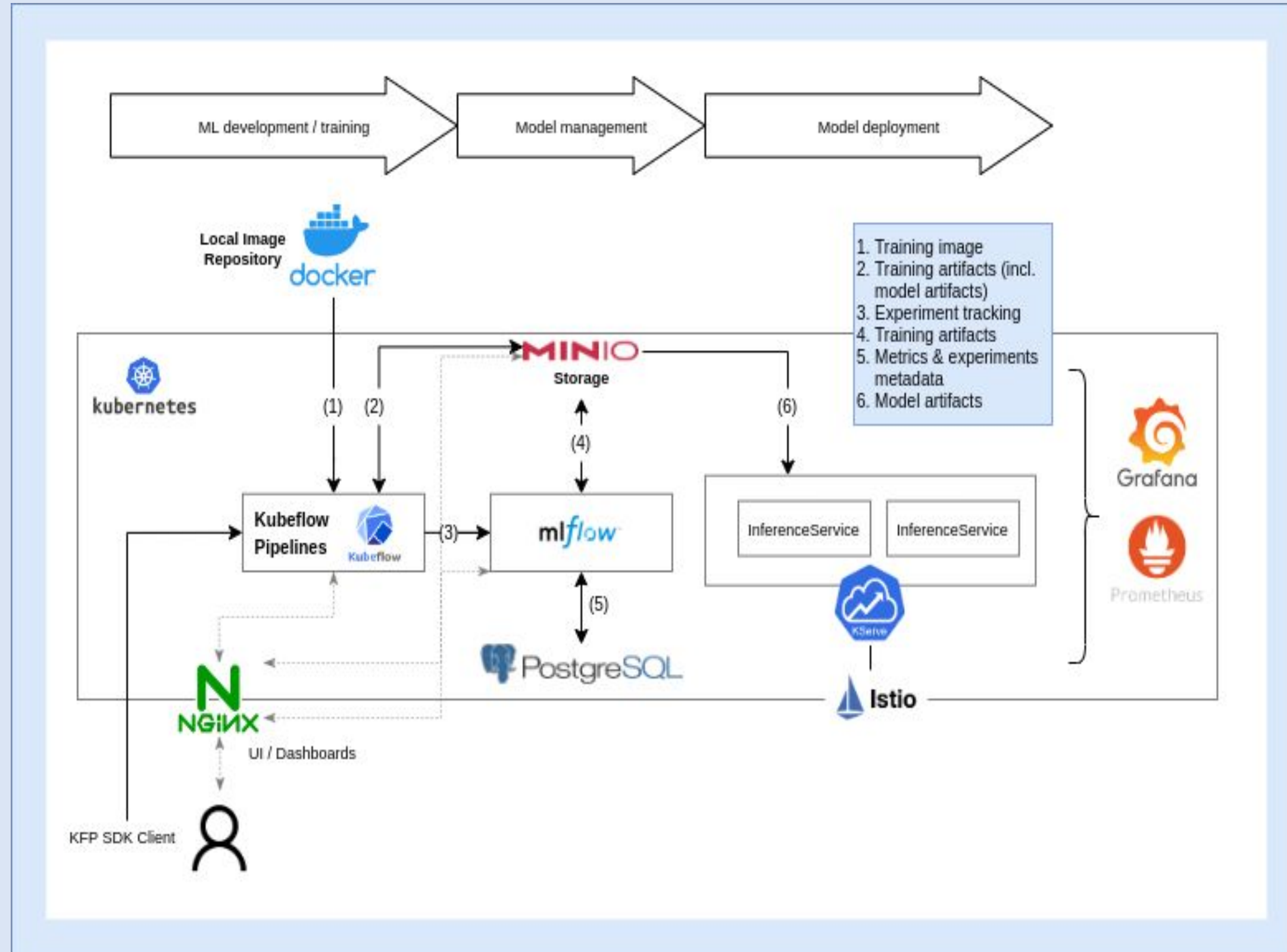


OSS MLOps created in IML4E is such a platform

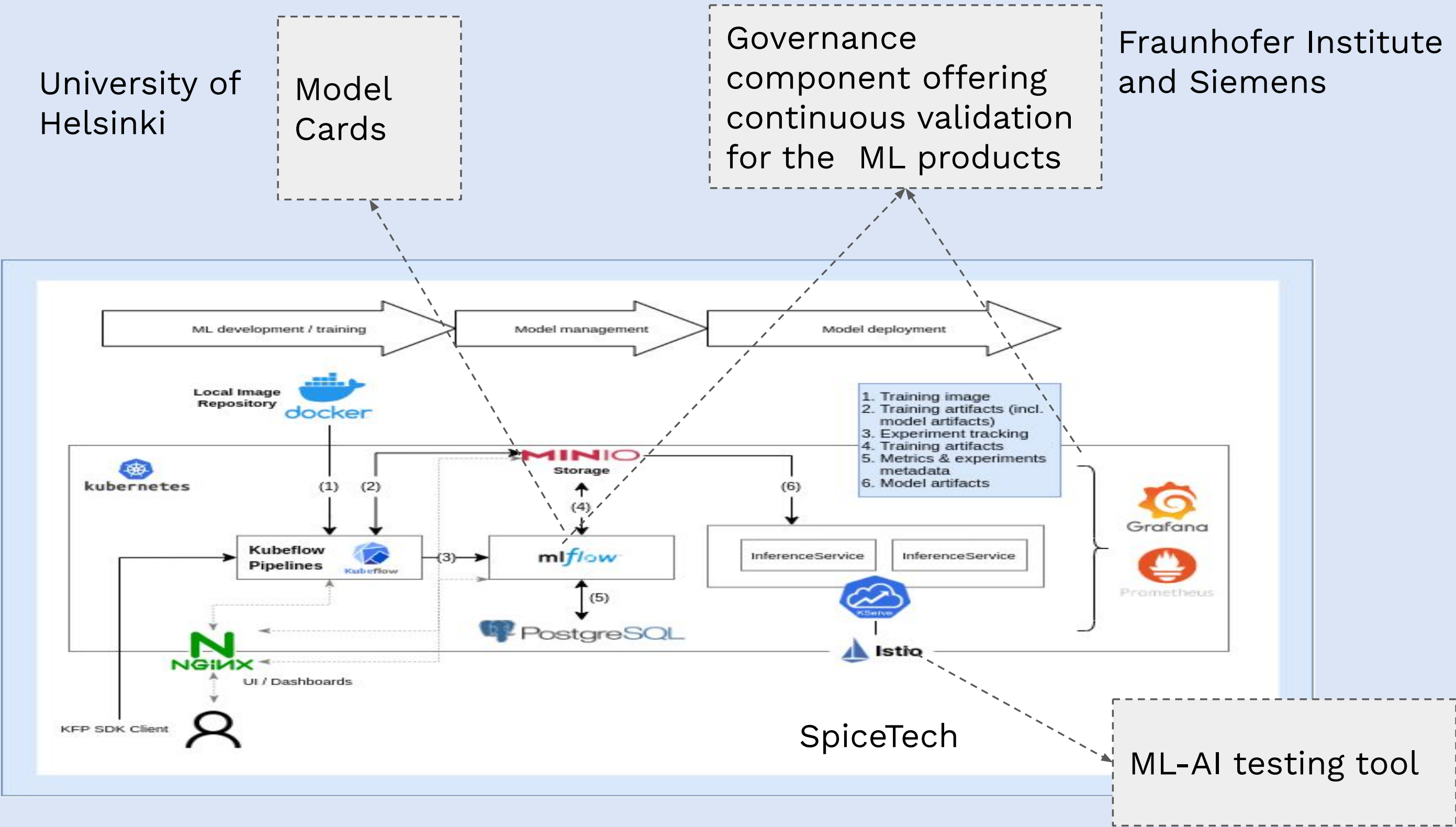
A platform up to date with latest ML/AI technologies.

Captures requirements in regards to

- automation,
- reproducibility,
- reusability
- scalability
- automation
- integrability



How OSS platform integrates with the rest of AI initiatives?



Technology choices in OSS MLOps

Industry-level and well-backed

- Encountered by Silo AI at many big and small clients
 - Telecom
 - Retail
 - Pharma
 - Automotive
 - ...
- Kubeflow (Google-backed, used in Google Cloud backends)
- MLflow (Databricks-backed, used in all clouds)
- KServe (part of Kubeflow ecosystem)

Not limited or limiting - extensible platform

- Added support for Ray later
- Added support for Slurm later



Unified AI development, deployment and monitoring experience across environments

1. AI professionals' own dedicated computer (local, lighter deployment)

- work laptop/desktop computer/virtual machine
- AI development and testing for individuals (developers, researchers, students)

2. AI teams and organizations common environments (full deployment)

- same workflows possible as on everybody's own computer
- usage in cloud (Google Cloud k8s, Azure and AWS VMs, others)
- usage in super-computing/HPC environments (Finnish IT Center of Science - CSC)



Success story: Leveraging supercomputing/HPC environment using OSS MLOps platform with CSC services



OSS MLOps platform in the CSC services

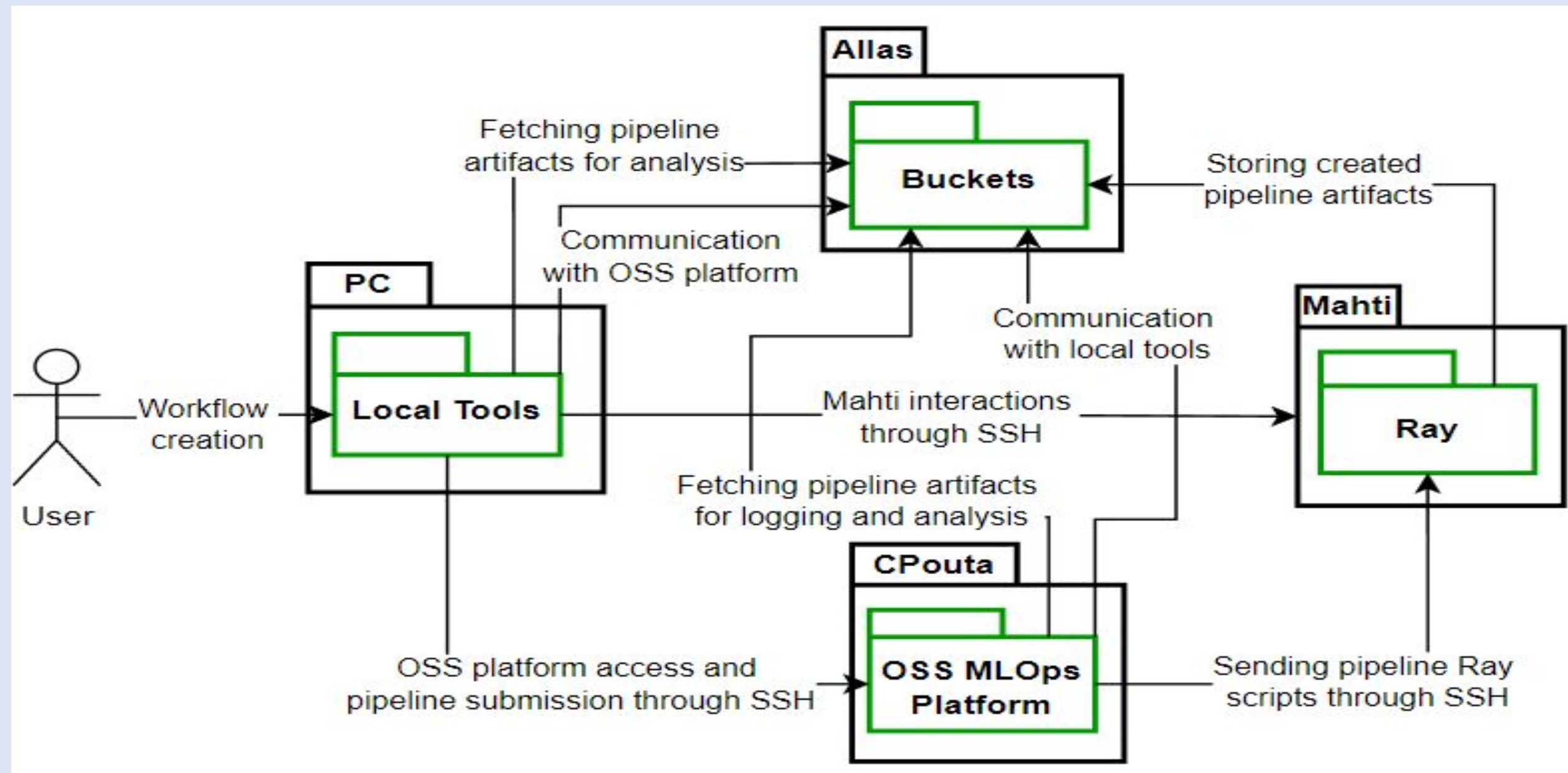
- OSS MLOps platform in the CPouta virtual machine service
- ML pipelines use Allas object storage and Mahti supercomputer for GPUs
- Customizable and scalable ML workflows through using Ray in Mahti through OSS MLOps

The platform could be used as a base for unified MLOps in CSC services

- Extend the support to CSC services EPouta (sensitive data) and Rahti (kubernetes)
- Extend the support to CSC services Puhti and LUMI (the fastest supercomputer in Europe, 5th in the world)
- Quantum platforms Kvasi and Helmi might enable quantum integration

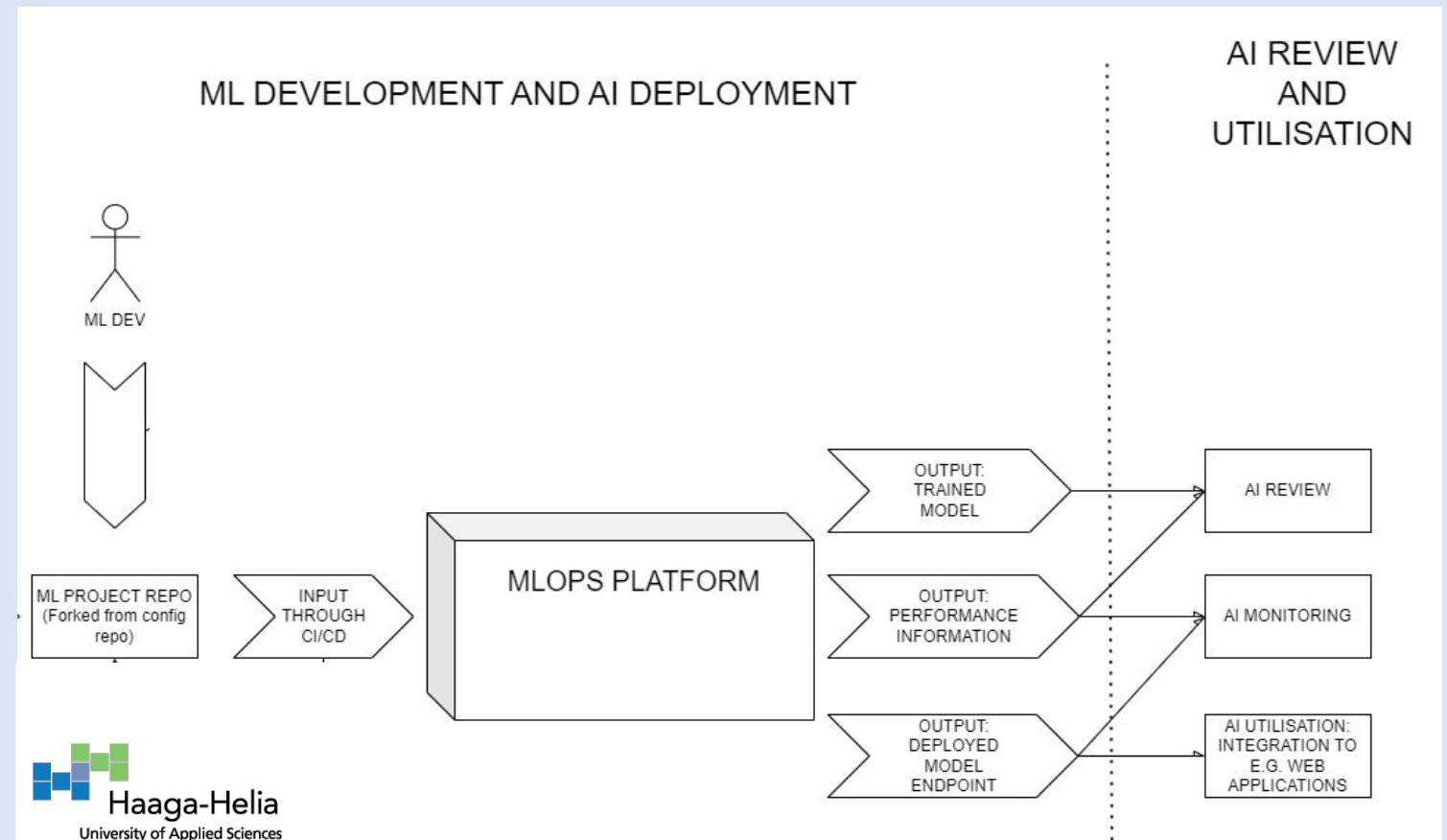


Architectural overview of cloud-HPC integrated OSS MLOps platform



Future directions in OSS MLOps development

1. Software DevOps (Git, CI/CD)
 - Modern AI/ML practices
2. Standardized use for HPC
 - Easy use at CSC for all Finnish researchers, students and collaborating companies
 - Support for LUMI?



Open source software MLOps platform (OSS MLOps)

Thank you!

Utilize freely without fees in your organization.

Join the users and/or the developers!

[GitHub - OSS-MLOPS-PLATFORM/oss-mlops-platform: OSS MLOps Platform](#)

