EUMaster4HPC

EUMaster4HPC Student Challenge

Kick - off meeting

September 18th 2025

EuroHPC Joint Undertaking

This project has received funding from the European High-Performance Computing Joint Undertaking under grant agreement No 101051997

Benchmarking AI Factories on MeluXina supercomputer

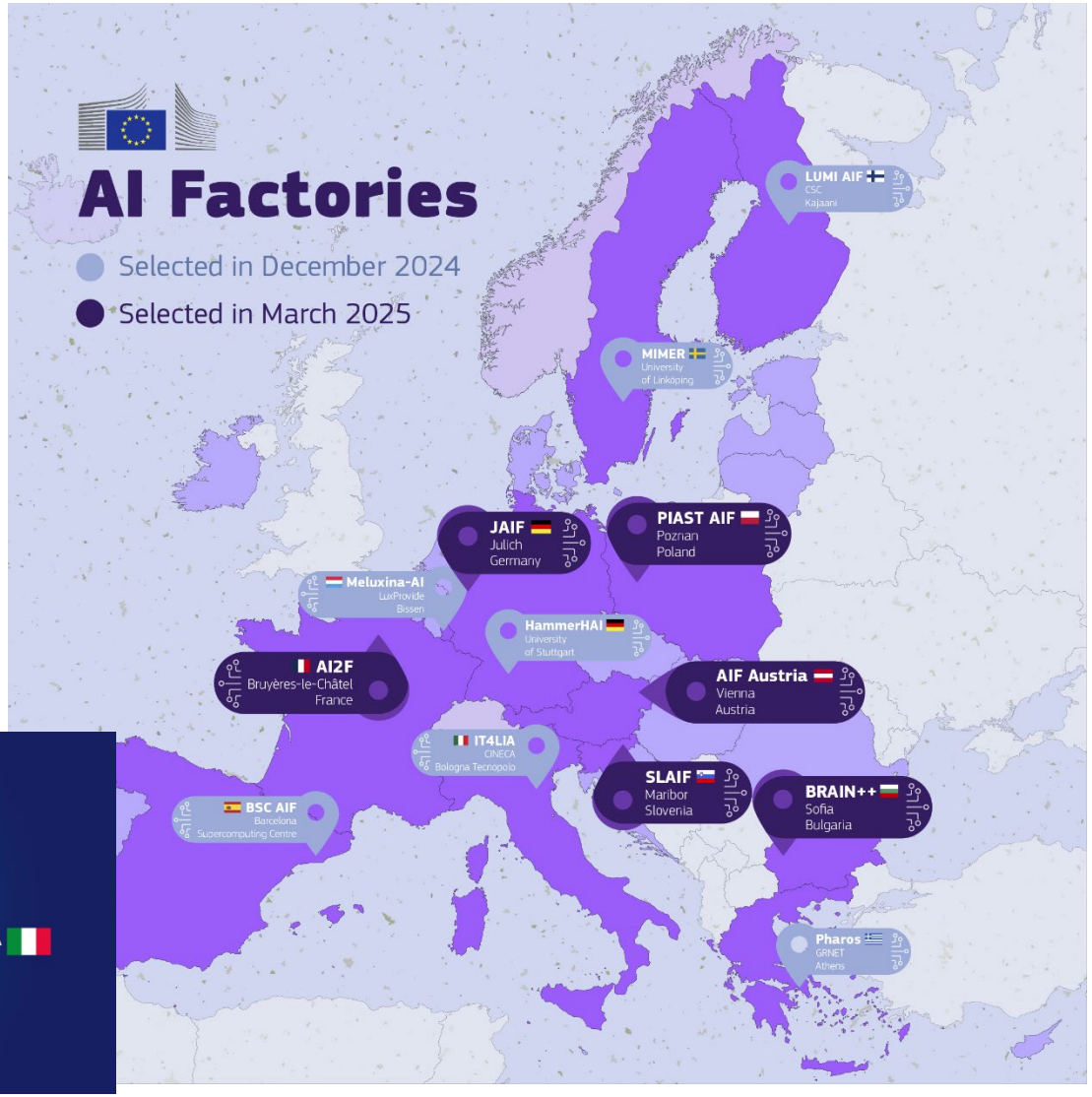
EUMaster4HPC Challenge 2025-2026

Dr Farouk Mansouri
Senior Solution Engineer
Farouk.mansouri@lxp.lu

Introduction – Why this Challenge?

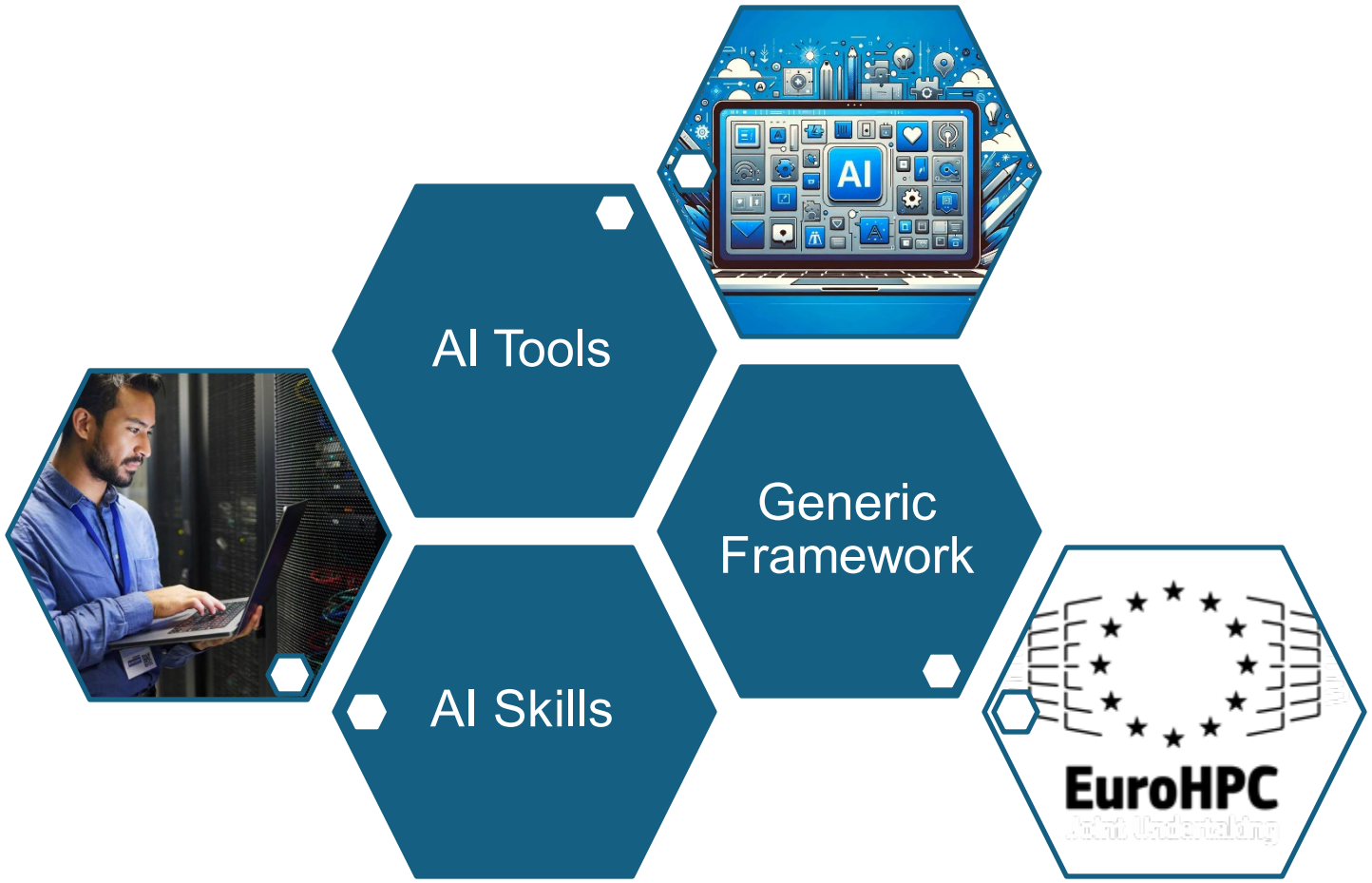
- AI Factories are coming to the EU: a mix of HPC, AI, and Data systems
- Growing need for scalable, reproducible AI infrastructure

This challenge empowers students to build tools for the next-generation AI Factory



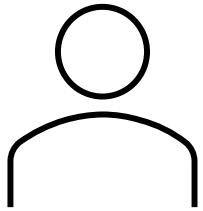
Challenge Global Objectives

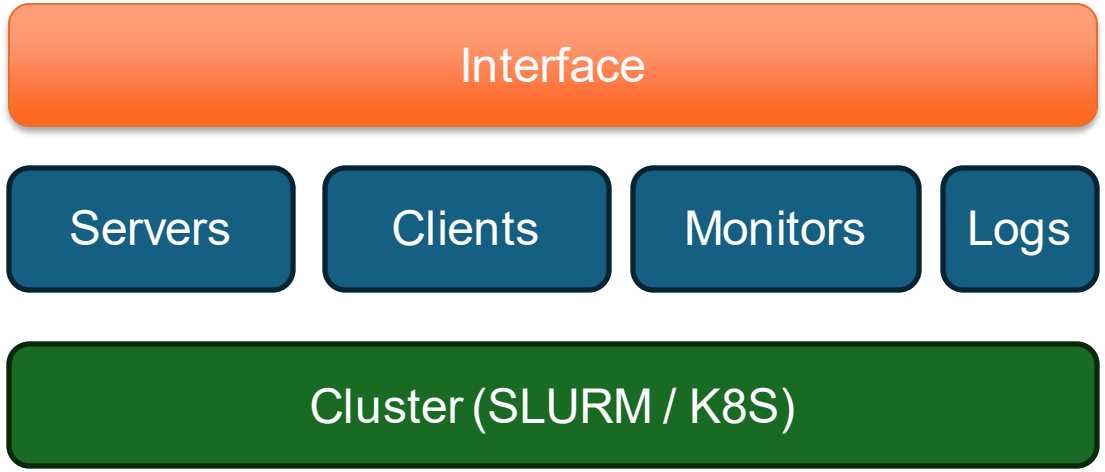
- Equip students with practical HPC + AI Development and deployment experience
- Encourage hands-on collaboration in a real HPC production environment
- Develop a framework to evaluate performance of AI Factory components widely used (Research & Commercial)
- Produce reusable benchmarking tools, insights, and dashboards



Challenge Architecture

Unified Benchmarking Framework for AI Workflows:





Servers:

- **Storage systems: File, Object, and Relational DBs**
- **Inference engines: vLLM, Triton**
- **Retrieval systems: Vector databases (Chroma, Faiss, etc.)**

Clients:

- **Large scale usage based on Slurm or K8S + scalable tools (Dask, Spark, …etc)**

Monitoring:

- **Monitoring: Prometheus + GrafanaOrchestration:**

Reporting:

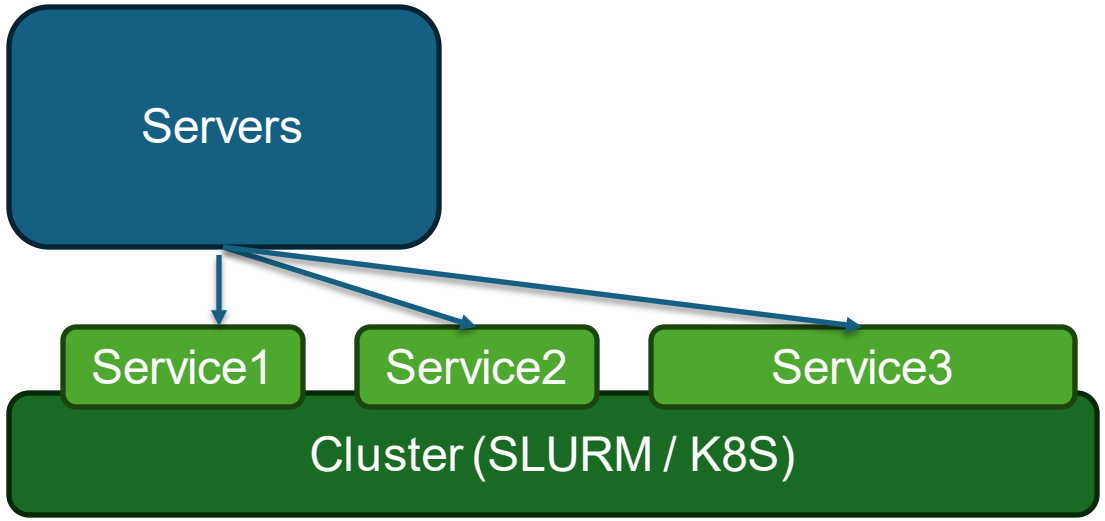
- **Diagram & Dashboards showing all experiments**

Interface:

- **APIs & GUI & CLI : Any kind of interface the user can manipulate to control the framework**

Challenge Architecture

Unified Benchmarking Framework for AI Workflows:



Servers:

It is a module to start services we want to benchmark with the following capabilities :

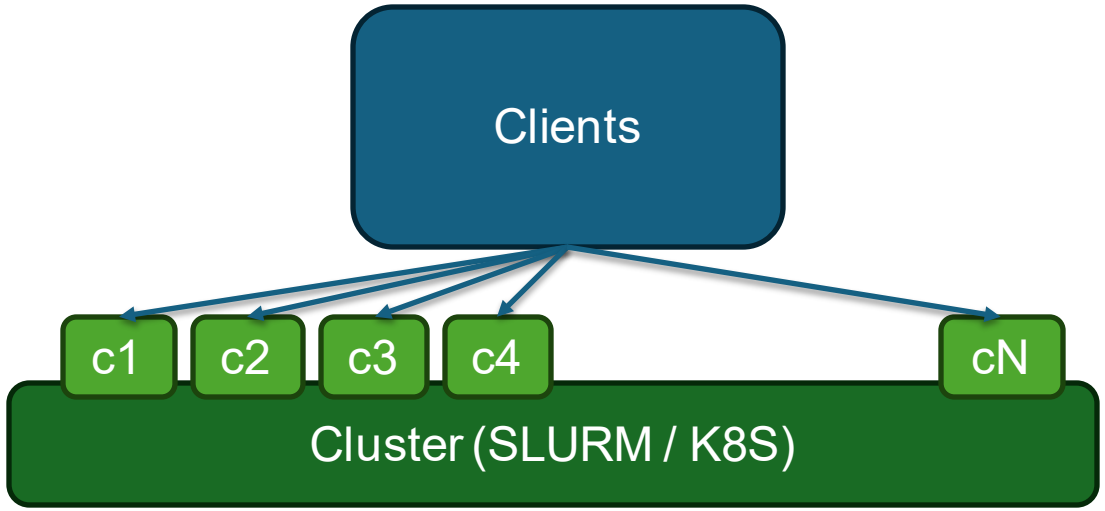
- **Server can be 1 node or several node execution**
- **Start one (or several) service(s) on HPC/K8S**
- **Stop one (or several) service(s)**
- **List available services (recipes)**
- **List running services**
- **Check service**

The examples we have are :

- **Storage systems: File, Object, and Relational DBs**
- **Inference engines: vLLM, Triton**
- **Retrieval systems: Vector databases (Chroma, Faiss, etc.)**

Challenge Architecture

Unified Benchmarking Framework for AI Workflows:



Clients:

It is a module to start a bunch on client to test one server service with the following capabilities :

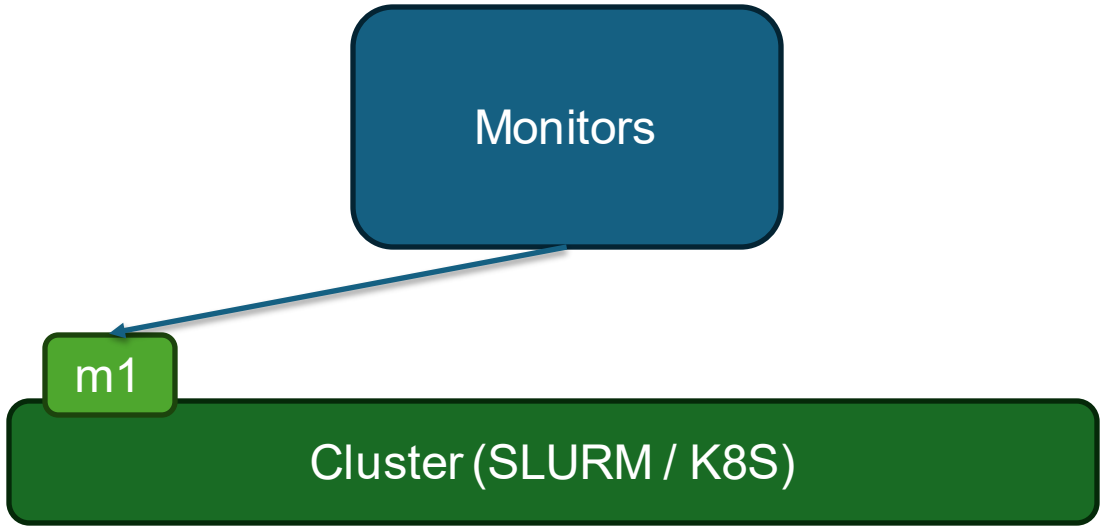
- **Clients can be single or multi nodes**
- **Start clients on HPC/K8S**
- **Stop clients**
- **List available client (recipes)**
- **List running clients**
- **Check client status**

The examples we have are :

- **Storage systems: File, Object, and Relational DBs**
- **Inference engines: vLLM, Triton**
- **Retrieval systems: Vector databases (Chroma, Faiss, etc.)**

Challenge Architecture

Unified Benchmarking Framework for AI Workflows:



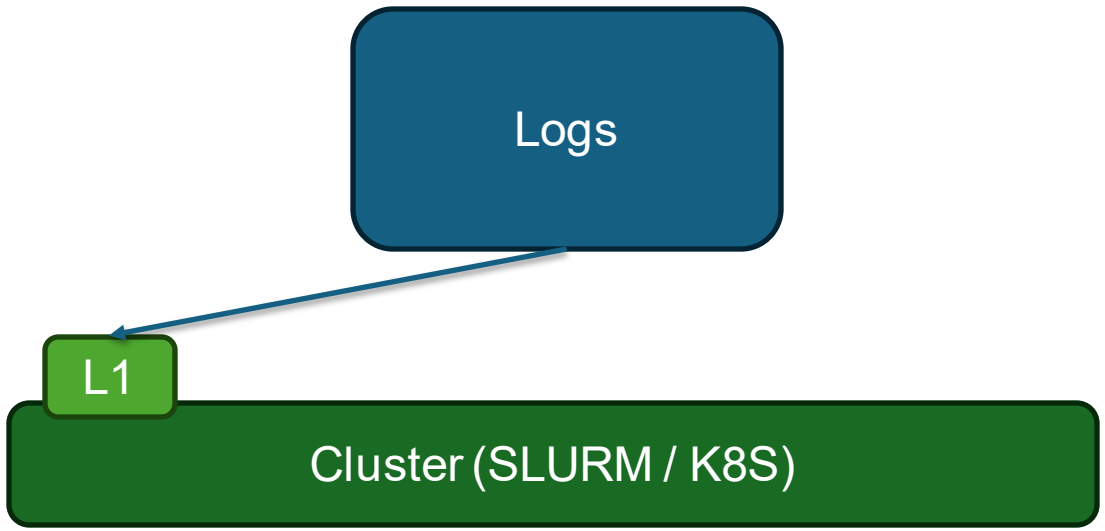
Monitors:

It is a module to monitor servers and clients started on the machine with the following capabilities :

- **Monitor instance collect server's metrics as described in the receipe**
- **Start monitor instance**
- **Stop monitor instance**
- **List monitor description**
- **List running monitor**
- **Check monitor**
- **Collect metrics in a file**
- **Show metrics**
- **Construct report**

Challenge Architecture

Unified Benchmarking Framework for AI Workflows:



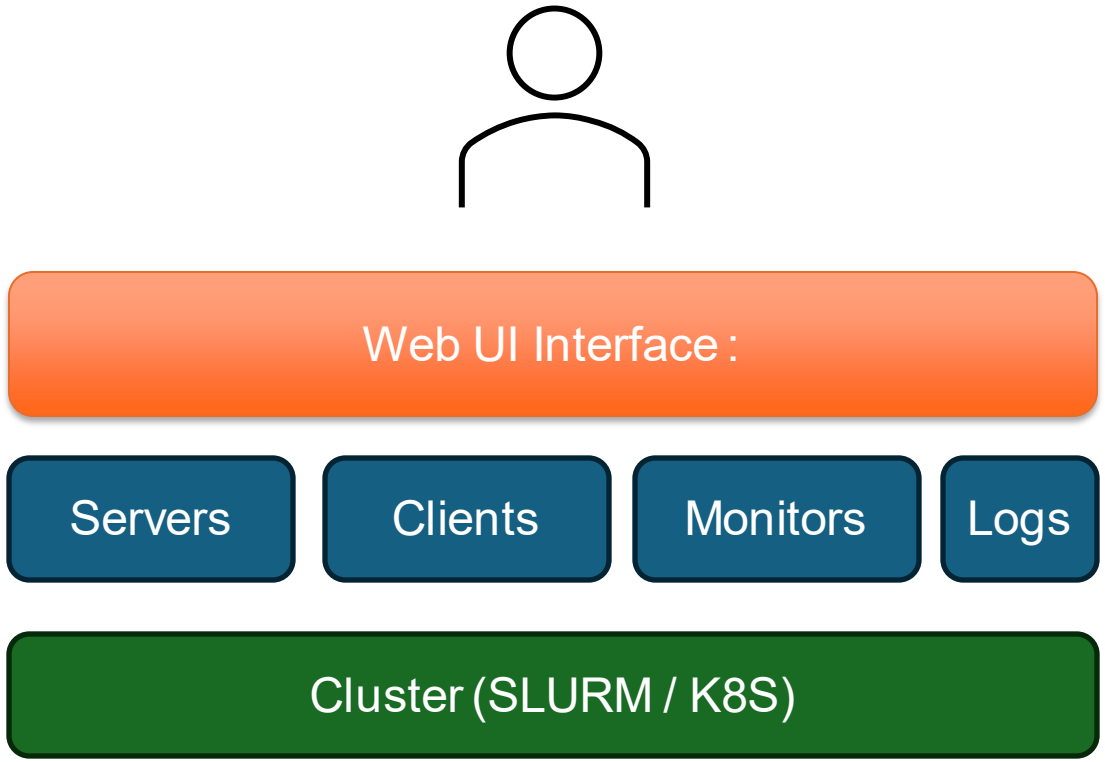
Logs:

It is a module to collect logs of servers and clients started on the machine with the following capabilities :

- **Monitor instance collect logs as described in the recipe**
- **Start log instance**
- **Stop log instance**
- **List logs**
- **Get logs**
- **Show logs**
- **Save logs**

Challenge Architecture

Unified Benchmarking Framework for AI Workflows:



Interface:

It is a module to manage at one place a benchmark experiment according to the recipe

- **Read and validate the recipe**
- **Start a benchmark session**
- **Stop a benchmark session**
- **List available bench recipes**
- **Show servers status**
- **Show client status**
- **Show logs**
- **Show metrics**
- **Save report**

Workflow Timeline (4 Months)

Phase	Activities	Delivery	Evaluation
Month 1	Onboarding, Exploration, Requirements definitions, Designs, Modules, Tasks	Github (README, Individual SLURM-Python examples, Issues)	Eval 1 : Review
Month 2	Modules development (Servers, Clients, Monitors, Reporters, Interface)	GitHub (Code : Modules, Tests)	Eval 2 : Review & Test
Month 3	Benchmarking Experiments, data collection, Raw results	GitHub (Results, Logs)	Eval 3 : Review and compare
Month 4	Evaluation, comparison, reporting, and defense	GitHub (Reports, Slides)	Eval 4 : Defense

→ Global hybrid meetings every month:

Mid-term meeting 1	October 2nd, 2pm
Mid-term meeting 2	November 4th, 2pm
Mid-term meeting 3	December 2nd, 2pm
Final meeting	January 12th, 2pm

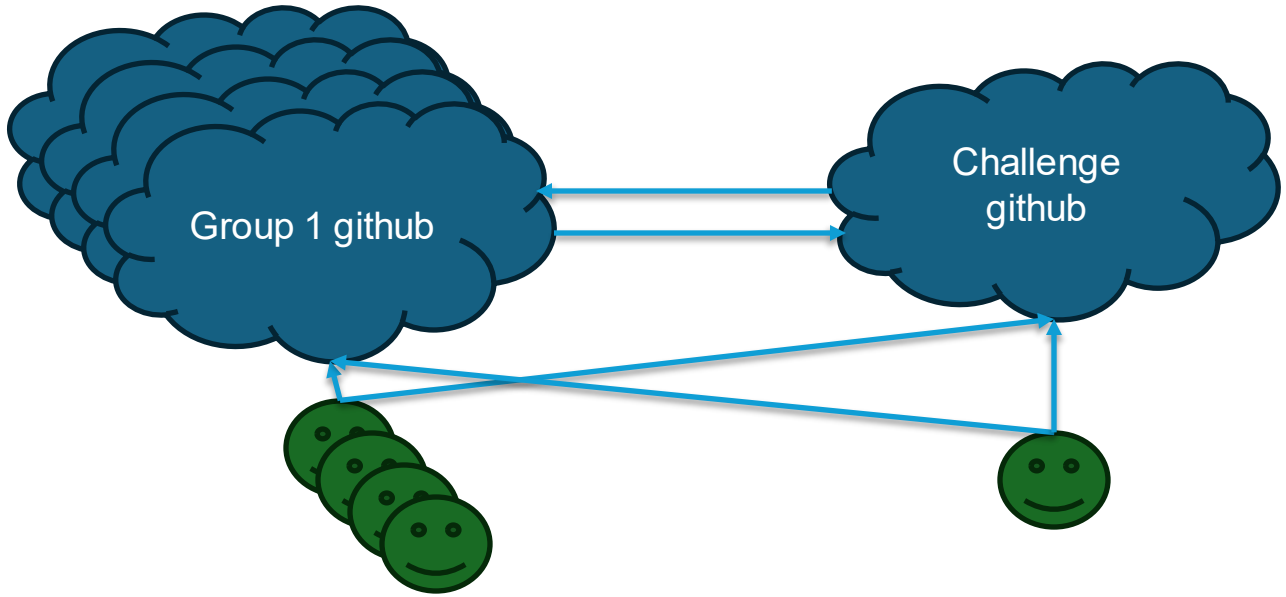
→ Biweekly technical meeting with the challenge supervisor to monitor each team's progress

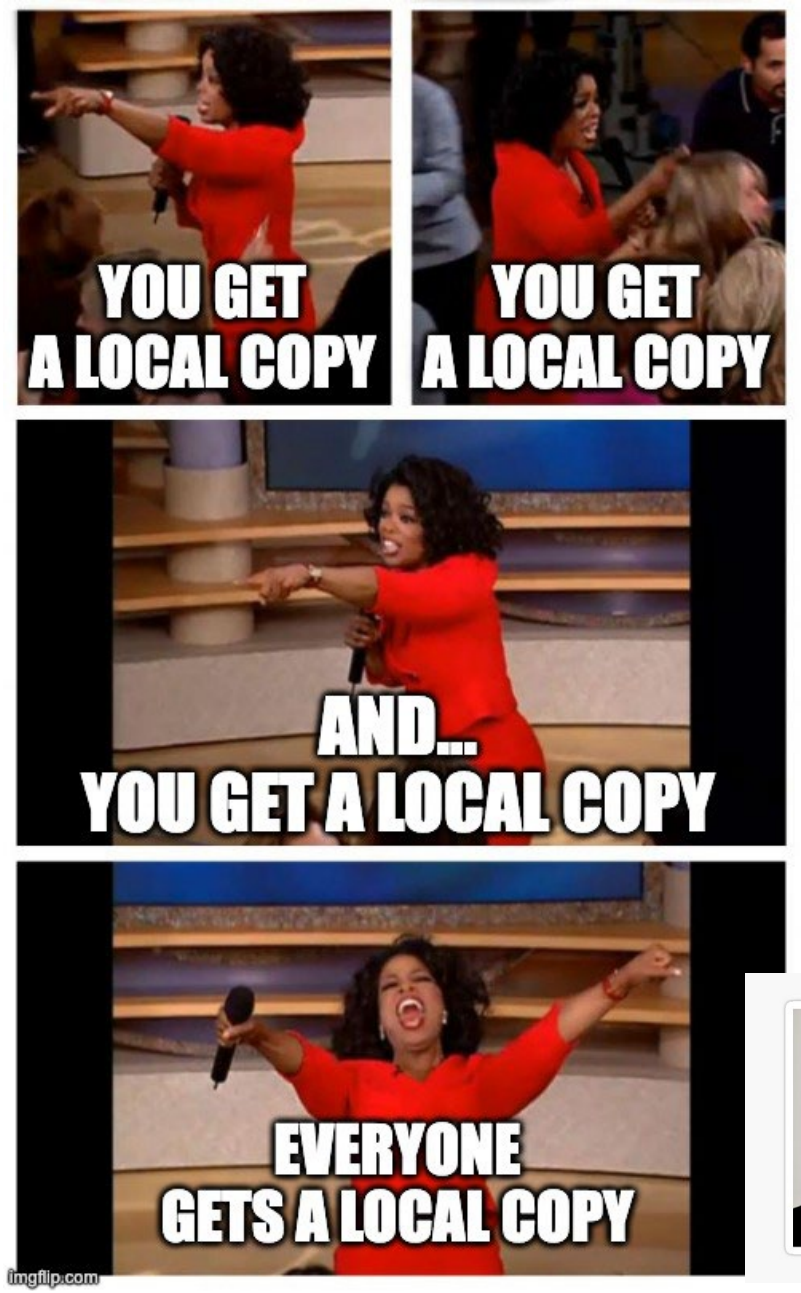
Space of work & Deliveries

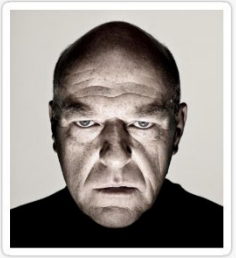
- Github source of thought : https://github.com/LuxProvide/EUMASTER4HPC2526

To follow the work, here are the steps :

- Phase 1: GitLab work
- Phase 2: GitLab work
- Phase 3: Gitlab work
- Phase 4: GitLab work

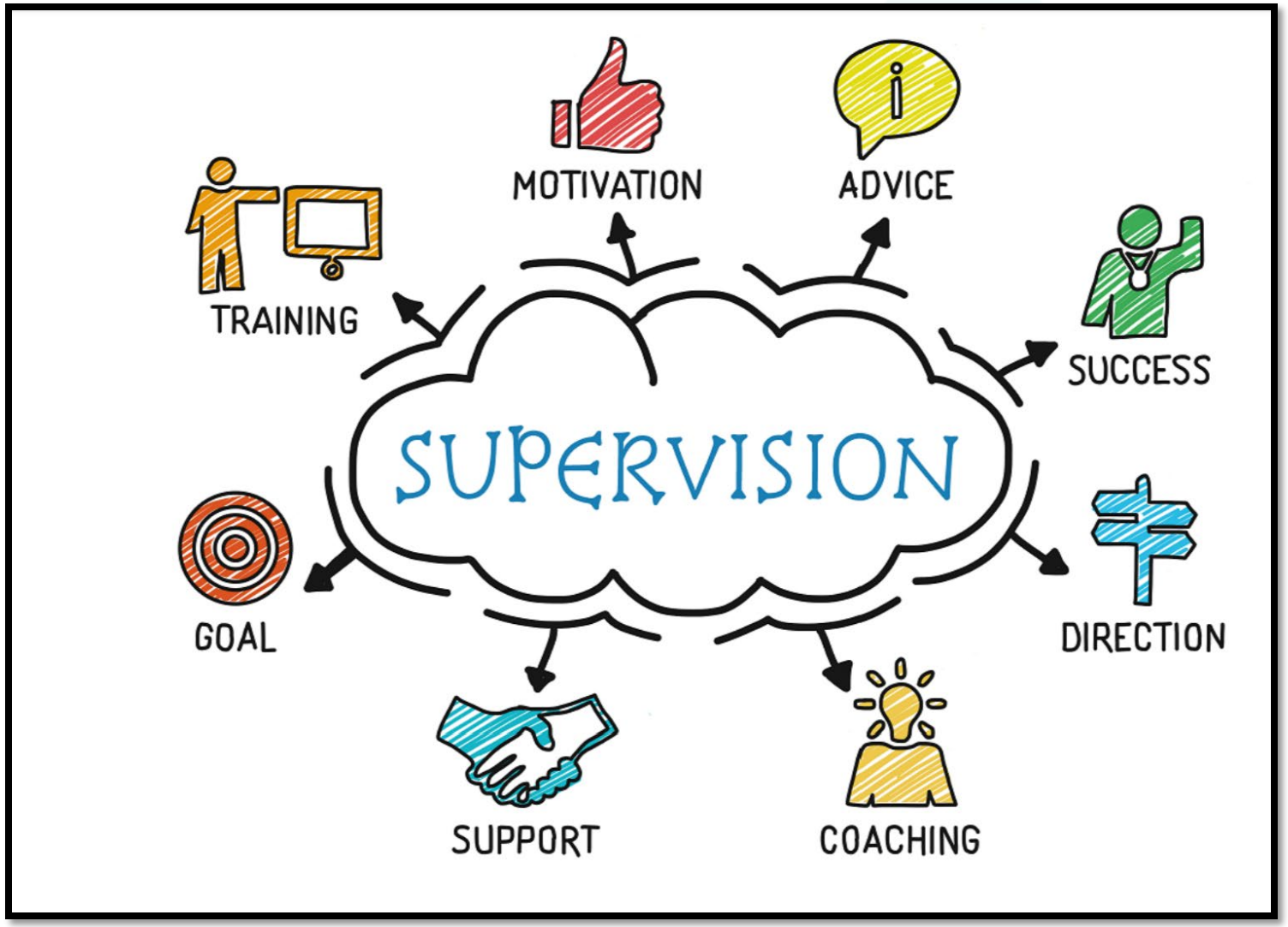






Supervision & Mentoring

- Dr. Farouk Mansouri – Challenge lead
- 1 onboarding meeting (Methodology, Project approach, Tech approach, …etc)
- 1 Eval meeting each month (End of month)
- Supervision, Methodology, Review, Evaluation
- LuxProvide Mentoring
- Weekly check-ins
- Guidance on architecture, Guidance on dev, benchmarking experiments, and deployment
- Group based unlocking and supporting (Q/A)



Expected Outcomes

- **GitHub repo with content : Design, Code, Issues, Results, Reports**
- **Reusable benchmarking toolkit for generic tools (use cases : storage, inference, and vector search)**
- **Real-time monitoring for performance visualization**
- **Reporting & Comparative analysis of AI workloads on MeluXina**

Students ready to operate within EU AI Factories

Team’s composition

Team 1

Thies Weel
Mario Capodanno
Can Beydogan
Giuseppe Galardi

Team 2

Andrej Cop
Tommaso Crippa
Dimitar Penkov
Alesio Demiri

Team 3

Emmanuele Caruso
Tsvetelin Tsetskov
Edoardo Leali
Sefa Böyükdikmen

Team 4

Giacomo Pauletti
Xavier l'Heureux
Jan Esquivel Marxen
Vittorio Cozzoli

Team 5

Alberto Taddei
Elizabet Koleva
Dennys Huber
Arianna Amadini

Team 6

Valerio Grillo
Michal Sterzel
Alessandro Ruzza
Davide Villani

Team 7

Luca Leonzio
Mihkel Tiks
Leonardo Evi
Jonah Holtmann

Team 8

Giulia Leonetti
Giovanni La Gioia
Laura Paxton
Alberto Finardi

Team 9

Daniele Kopyshevskiy
Mohamed Mandour
Nicola Noventa
Ivan Al Khayat

Team 10

Filippo Wang
Matteo Arrigo
Leon Ackermann
Christian Karg

Team 11

Emanuele Lovino
Luca Lamperti
Thomas Gantz
Patrick Cerka

Thank you

Follow the project on Social Media

EuroHPC Joint Undertaking

This project has received funding from the European High-Performance Computing Joit Undertaking under grant agreement No 101051997