AI-DAYS@HES-SO 2025 – GENEVA & LAUSANNE – JANUARY 27- JANUARY 29, 2025

WORKSHOP DAY JANUARY 27

Session 1: Compute Infrastructures for IA applications in the wild

With the advent of Chatbots, LLMs and other generative IA technologies, as well as other progresses in the IA field, there is an explosion of the demand for compute force. IA is no longer computer science: it is computational science. As such, it can no longer be done with casual, self-managed equipment. More advanced compute infrastructures are required both to satisfy user needs (in terms of compute power, GPU Ram capacity) and to ensure a decent utilization of the increasingly costly resources.

Content and topics

The purpose of this workshop is to gather people in charge of compute infrastructure (on-prem, cloud or hybrid) destined to support AI workloads (both training and inference). Being "in charge" means someone who does one or more of the tasks below

- Prepares and follows purchase orders for servers including GPUs.
- Racks the servers, ensure they are connected with adequate network performance
- Install the OS, the drivers, some software stack
- Binds the servers with the login system, typically LDAP
- Monitor the utilization of the server over time
- Ensure there are no abuse
- Insert the server into a cluster (e.g. SLURM, K8s)
- Designs and implement cluster architecture
- Helps AI engineers to run their workloads according to the policies
- Manages cloud platforms access (AWS)

The workshop will be organized around keynote and invited presentations, but above all around **short presentations (~15min) during which participants will present one of the following** (not limited to) :

- A technology under test or underutilization in their group (e.g. code, Determined.ai, etc.)
- A picture of how workloads are mapped onto hardware in their group
- A particularly painful aspect of IA-infra there are experiencing in their group
- A project for improving the situation in their group

We also plan to run breakout groups who will explore ways to achieve better collaborations between institutions.

The workshop has no specific registration, and walk-ins are welcome.

Schedule: 8h30-12h + 13h-17h30

The workshop schedule is under construction. We are looking for **short presentations**! Please contact Sébastien if you have an interest in presenting.

Workshop committee

Last Name	First Name	Institution	e-mail address
Rumley	Sébastien	HEIA-FR	sebastien.rumley@hefr.ch
Gambin	Dorian	HEIG-VD	dorian.gambin@heig-vd.ch
Stadelmann	Marc	ZHAW	marcandre.stadelmann@zhaw.ch
Kucharavy	Andrei	HE-VS	andrei.kucharavy@hevs.ch
Menetrey	Jämes	UNINE	james.menetrey@unine.ch
Marques Reis	Henrique	HE-Arc	henrique.marquesreis@he-arc.ch

swiss ai center

Swiss AI Center for SMEs

CSIA-PME – Centre Suisse d'Intelligence Artificielle pour les PMEs