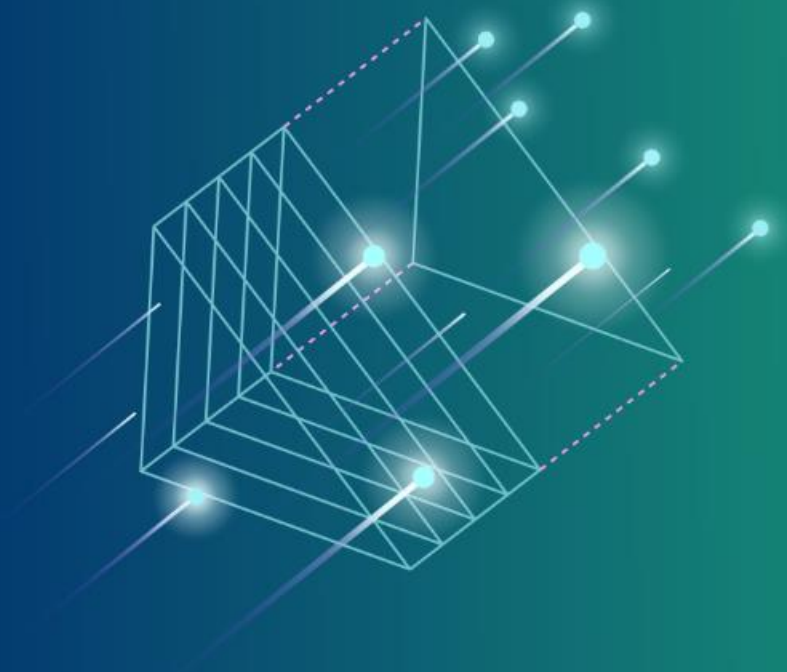
AI Factories and the data challenge

EUROPEAN
BIG DATA
VALUE FORUM
2-4 OCT | BUDAPEST - HUNGARY



EUROPE
FOR GLOBAL
LEADERSHIP
IN AI & DATA

ORGANISED BY
BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH
European Commission
ELTE EÖTVÖS LORÁND UNIVERSITY
IDEAL-IST
NTP Neumann Technológiai Platform
HUN REN | SZTAKI
AI & AUT EXPO

BDVA motivation and support

Thomas Hahn

President BDVA
Chief Expert Software at Siemens AG

Speakers

- Juan Pelegrin (European Commission CNECT C2 HPC unit)
- Aleksi Kallio (Development Manager, CSC – IT Center for Science)
- Sergi Girona (Operations Director, Barcelona Supercomputing Center)
- Georg Rehm (Principal Researcher, DFKI)
- Lilit Axner (EuroHPC JU)
- Laura Morselli (Solution Architect @ HPC, CINECA)
- Branislav Jansik (Supercomputing Services Director, IT4Innovation / VSB)
- Mike Matton (Innovation projects and partnerships coordinator, VRT)
- Andrejs Vasiļjevs (Executive Chairman, Tilde)

Agenda

View from policy makers			
14:05 – 14:20	15 min	• Overview of the AI factories, and the specific data requirements	• Juan Pelegrin (European Commission CNECT C2 HPC unit) • Lilit Axner (EuroHPC JU)
Panel discussion with AI Factories potential hosting entities			
14:20 – 14:55	35 min	• Challenges • Data needs • Expectations	• Aleksi Kallio (CSC – IT Center for Science) • Sergi Girona (Barcelona Supercomputing Center) • Branislav Jansik (IT4Innovation / VSB) • Laura Morselli (CINECA)
Panel discussion with JU, data spaces and potential users			
14:55 – 14:30	35 min	• Potential scenarios • Data offer • User needs • Collaborations	• Lilit Axner (EuroHPC JU) • Georg Rehm (DFKI, Language Data Space) • Andrejs Vasiļjevs (Tilde) • Media (TBD)

European
Commission
AI Factories
2/3 September 2024

AI Factories in President von der Leyen's new political guidelines

*"Through our **Artificial Intelligence (AI)**, Europe is already leading the way on making AI safer and more trustworthy, and on tackling the risks stemming from its misuse.*

We must now focus our efforts on becoming a global leader in AI innovation.

*In the first 100 days, we will ensure access to new, tailored supercomputing capacity for AI start-ups and industry through an **AI Factories initiative**.*

*We will also develop with Member States, industry and civil society an **Apply AI Strategy** to boost new industrial uses of AI and to improve the delivery of a variety of public services, such as healthcare."*

Ursula von der Leyen, 18 July 2024

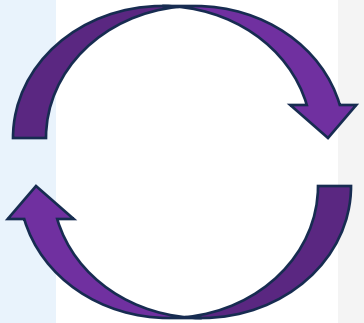
POLITICAL GUIDELINES FOR THE NEXT EUROPEAN COMMISSION 2024–2029

European Commission

AI in the EU: AI Act and AI Innovation Package

The 2024 AI Act

- The **world's first** comprehensive AI law
- A **horizontal** act with a **risk-based** approach – different rules for different risk levels
- Protection of **health, safety** and **fundamental rights**
- **Innovation friendly**
- Will apply to **public and private** actors, **inside and outside the EU** (as long as the AI system or general-purpose AI model is placed on the Union market), **providers and deployers**
- Sets up the **AI Office -** centre of AI expertise ; foundation of a **single European AI governance** system; key role in implementing AIA



The 2024 AI Innovation Package

- **Boosting startups and innovation in trustworthy AI**
- **Two main components**

(1) AI Factories:

Making available HPC computing capacity to facilitate the development of GenAI models/ Applications

(2) AI support measures with GenAI4EU:

Stimulating the development in strategic sectors of novel and innovative applications based on GenAI models and facilitating their uptake

AI Factories Components

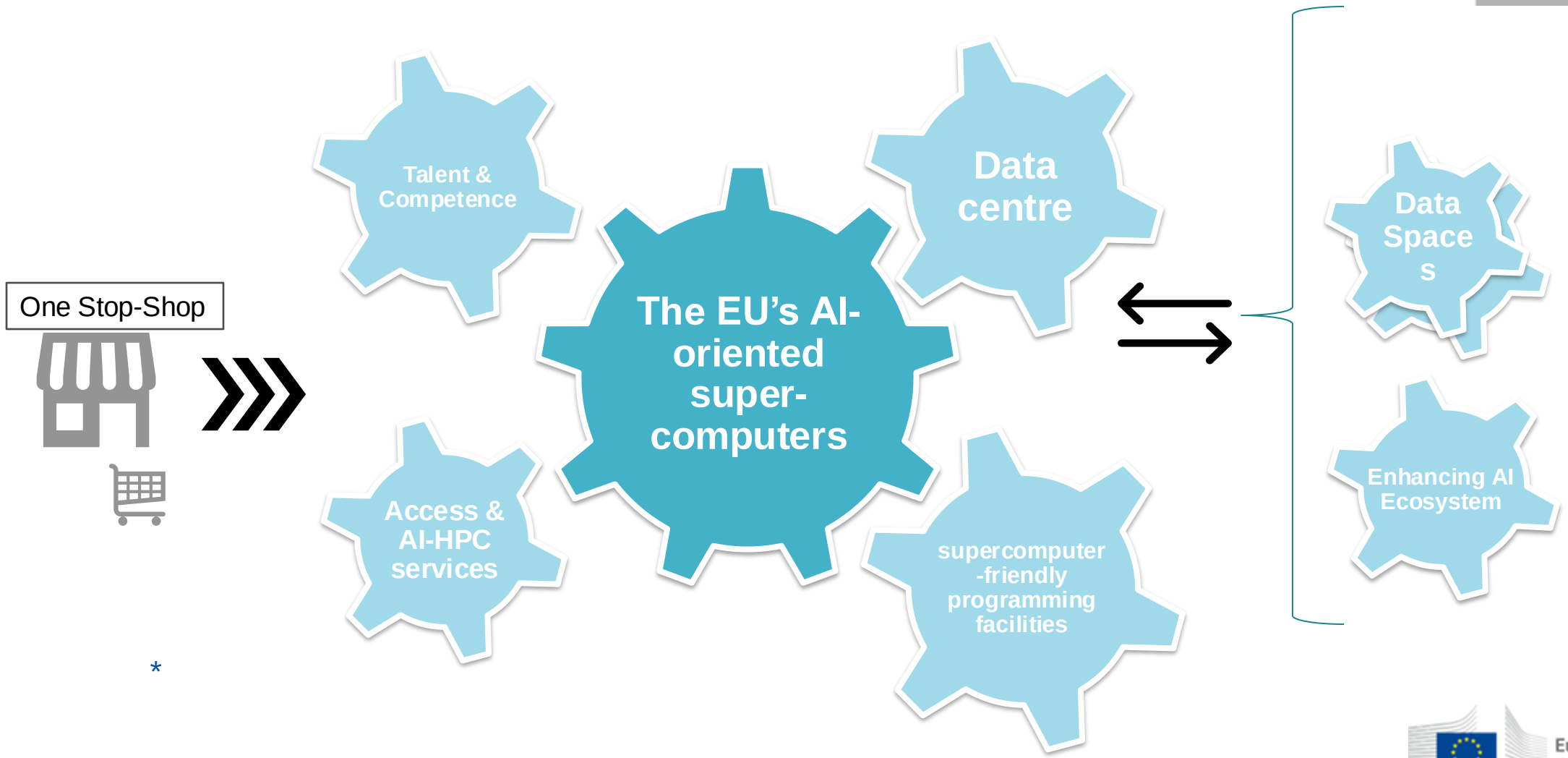
AI-optimised Supercomputer

+

Associated “AI Factories” activities and services

AI Factories will provide the **computing power, data, support and talent**, to offer a wide and exhaustive **range of services to AI startups and researchers** needed for the development of **European generative AI models and applications**

AI Factories: Empowering the AI Ecosystem



*

AI Factories and their Services

EU strategy on AI Factories

(Networked) AI Ecosystems

- **AI supercomputers (new)**
- **Upgrades of (AI-) supercomputers**
- **Dedicated HPC/AI services**
- Application support
- Access policy to supercomputers
- **Access to data / common EU data spaces**
- Support to AI EDIC (ALT-EDIC)
- Human talent and skills
- Collaboration with AI Office – **EU values**

National strategy on AI Factories

Invest in an AI ecosystem

- ❑ AI-supercomputer
- ❑ Data centre(s)
- ❑ Access to data + open gov. data
- ❑ Dedicated services
- ❑ Human talent and skills, incl. investing in housing facilities
- ❑ Cooperation with Universities
- ❑ Local GPU clusters
- ❑ Digital Innovation Hubs
- ❑ AI start-up policy (access to capital, tax incentives, etc.)

Data spaces | Access to supercomputers | HPC for AI services | TEFs for AI

Data centre | Talent | Cooperation with universities | Digital Innovation Hubs for AI

Activities of AI Factories (1/3)

- **Compute:** demonstrate a minimum available or planned AI-optimized computing capacity required for supporting an AI Factory ecosystem
- **Storage:** co-located with the supercomputer / high-speed / capacity storage
- **Connectivity:** High-bandwidth, low-latency secure networking
- **Software & Application Development:** Ready-to-use set of AI-oriented tools
- **Data Centre:** Data centre facility co-located with supercomputer

3. Activities of AI Factories (2/3)

- **User Support:** support activities to be offered to users (e.g., optimising the computing usage to speed up training, optimising the model for efficient deployment); 24/7 service
- **Skills / Education:** Skills Plan outlining the skills needed for the AI stakeholders (e.g. HPC advanced courses (deep learning, AI programming environments, hackathons, traineeships)
- **Engagement / Interacting with AI community:** Interacting with AI-ecosystem at large & other AI initiatives (e.g., TEFs, EDIHs, NCCs)
- **Guaranteeing trustworthy AI:** Collaboration with the AI office
- **AI Factories networking:** A collaborative framework to ensure an effective networking & resource optimisation among AI Factories

ALL ELEMENTS SHOULD BE ALIGNED AND MUTUALLY REINFORCE EACH OTHER

Activities of AI Factories (3/3)

- **Data:**
  - Data is at the core of Data Centric AI Gen Models
  - Availability and access to high-quality data repositories
  - Identification Knowledge Corpus
  - Identify access to selected Common European Data Spaces
  - Security: Protection of sensitive data
  - Secure Trusted Environments: results are usable & trustable (e.g. compliant with AI Act)

- Call to Action: Bridge the distance between AIF & Common Data Spaces

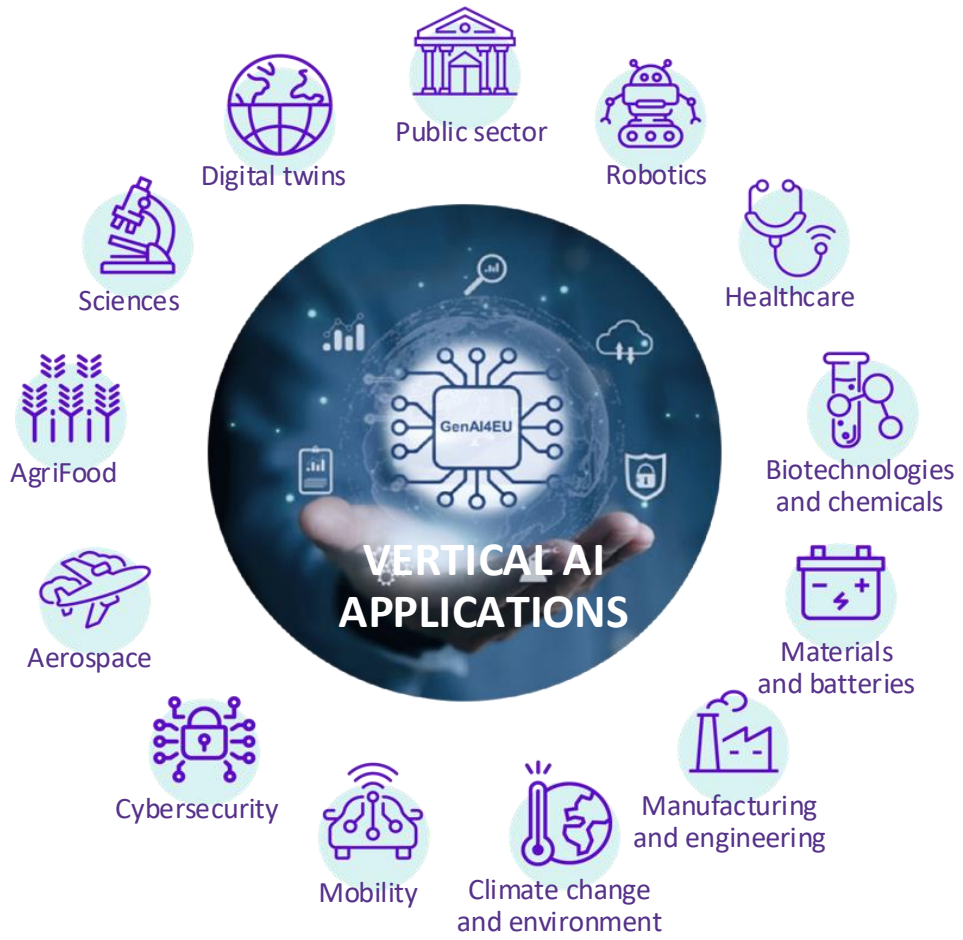
AI Factories: Strengthening the AI Ecosystem

STARTUPS

UNIVERSITIES

EUR 2 bn

EUR 500 million in applications (GeAI4EU initiative)

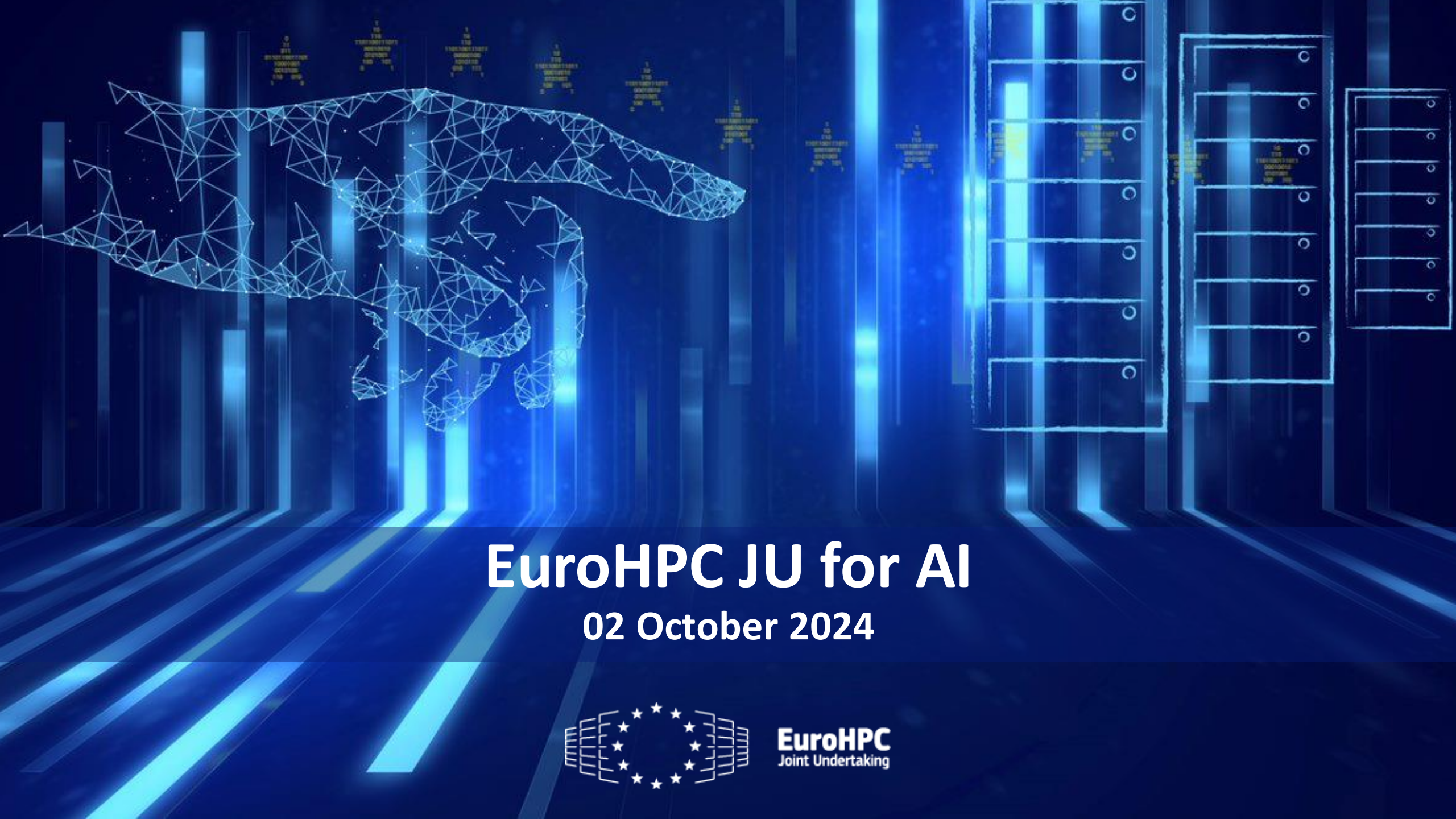


About 6300 AI Startups in the European Union

About 300 are generative AI Startups

Thank you!

European Commission



EuroHPC JU for AI

02 October 2024

EuroHPC
Joint Undertaking

EuroHPC JU Regulation

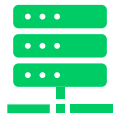
The EuroHPC JU mission is to establish a world-leading hyper-connected and federated HPC service and data infrastructure ecosystem in the Union.

AI = HPC + Expert Support + DATA

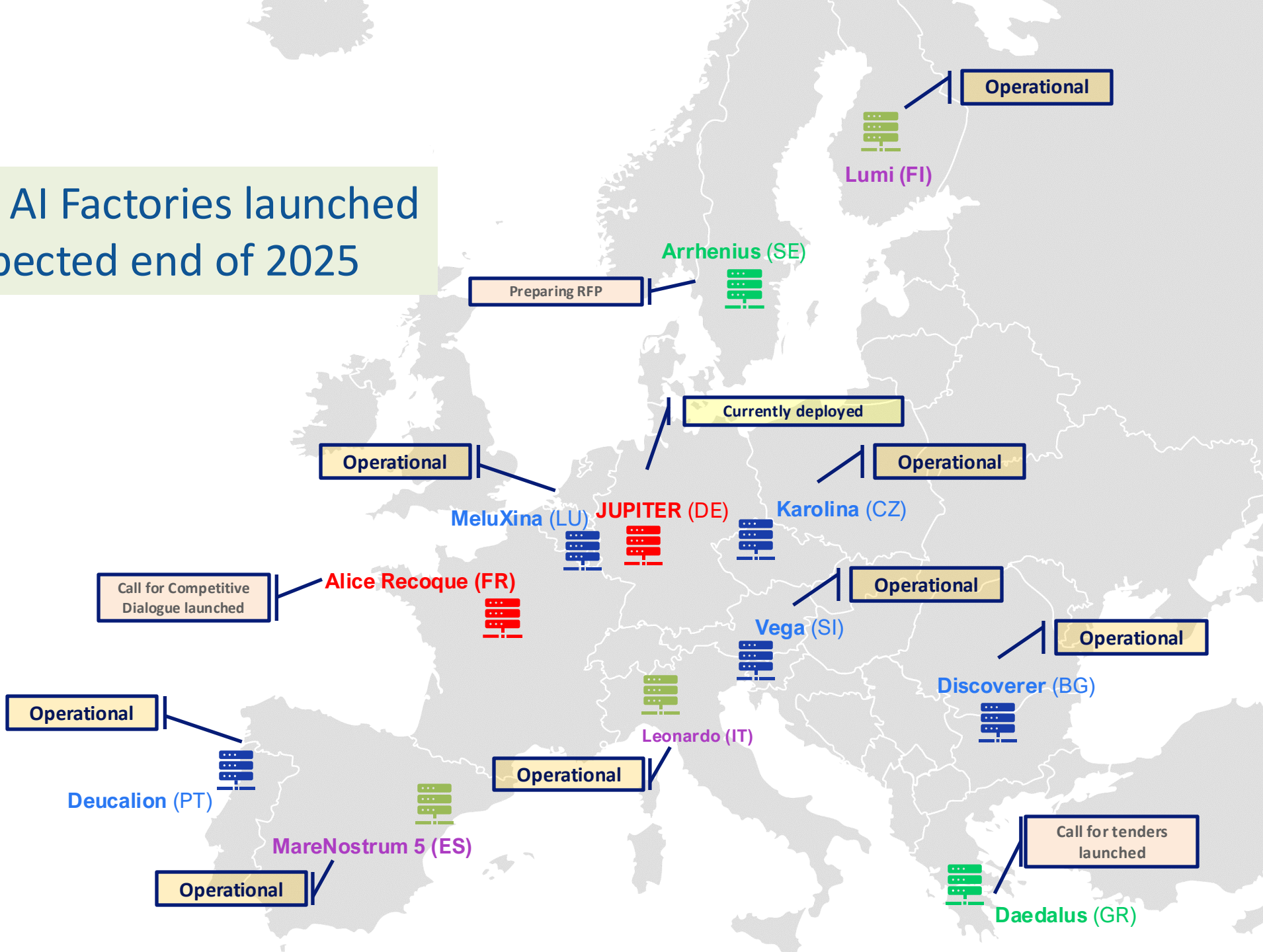
Call for selection of AI Factories launched
First systems expected end of 2025

EXASCALE

PRE-EXASCALE


MID-RANGE

PETASCALE



As of October 2024: EuroHPC JU HARDWARE for AI

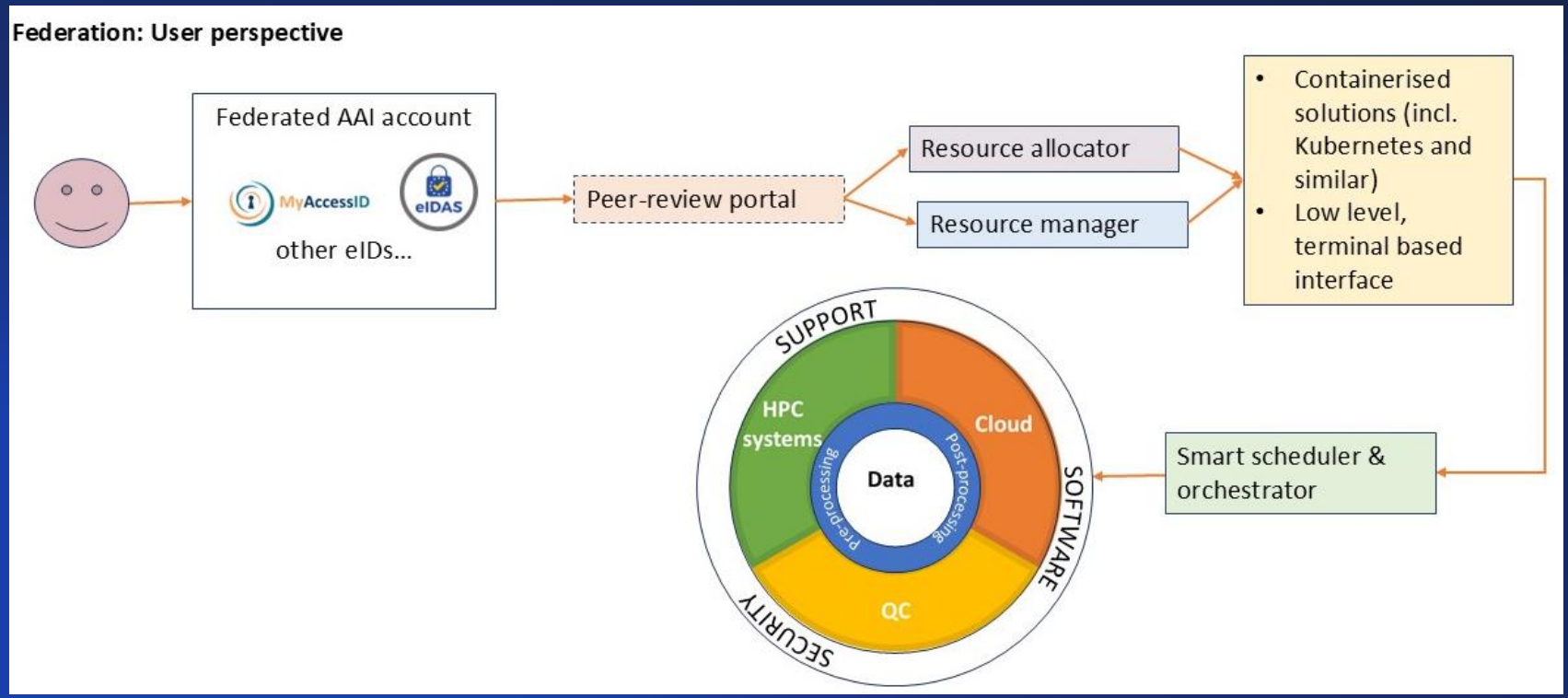
- ✓ **9** Supercomputers up and running (of which 3 are pre-exascale)
  - ✓ These supercomputers together have about **44 000 GPUs**
  - ✓ NVIDIA A100, NVIDIA H100, AMD MI250X
  - ✓ Only through EuroHPC JU calls over **20 000 000** aggregated **GPU Node (*) hours per year** are available for users (or on average more than **130 000 000** GPU hours)
  - ✓ The national parts of the supercomputers provide even more GPU hours through national calls of access

(*) the numbers to be multiplied by 4 or 8 based on the given supercomputer architecture

Federated and Hyperconnected

- ✓ Hyperconnectivity Study to be finished in 2024
  - ✓ Outcomes to be analysed to understand the Europes need of Hyperconnectivity

- ✓ EuroHPC Federation Platform: A Unified platform to Connect European HPC and Data infrastructures



As of October 2024: Euro

AI FACTORIES:

All under one umbrella and more …

- ✓ **AI Support Centre:** To t… of advanced HPC capabilities for large-sc… AI model…
- ✓ **FFplus or Fortissimo Pl**… …ogies such as HPC and AI to improve their… …eu/
  - ➢ Since 2020, 42 Exp…
- ✓ **EPICURE**: To provide se… …g AI, encompassing applicati…
  - ➢ Expert support
  - ➢ Training
- ✓ **33 EuroCC National Competence Centres:** To provide su…port to SMEs, large industries, the public sector and academia to adopt both HPC, and AI on HPC. https://ww…v.eurocc-access.eu/
  - ➢ Expert support
  - ➢ Training

THANK YOU

For more information, feel free to visit our website and social media:

/eurohpc-ju.europa.eu /EuroHPC_JU /eurohpc-ju /eurohpc-ju

EuroHPC Joint Undertaking

EUROPEAN BIG DATA VALUE FORUM

PANEL DISCUSSION WITH AI FACTORIES POTENTIAL HOSTING ENTITIES

Aleksi Kallio
Development Manager
CSC – IT Centre of Science Ltd

Sergi Girona
Operations Director
Barcelona Supercomputing Center

Branislav Jansik
Supercomputing Services Director
VSB Technical University of Ostrava

Laura Morselli
Solution Architect @ HPC,
CINECA

EUROPEAN BIG DATA VALUE FORUM

PANEL DISCUSSION WITH EUROHPC JU, DATA SPACES AND POTENTIAL USERS

Georg Rehm
Principal Researcher
DFKI

Andrejs Vasiļjevs
Executive Chairman
Tilde

Lilit Axner
Programme Officer Infrastructure
EuroHPC JU

Mike Matton
Innovation projects and partnerships coordinator
VRT

EUROPEAN BIG DATA VALUE FORUM
2-4 OCT | BUDAPEST - HUNGARY

ebdvf.eu
#EBDVF24

Thank you!

ORGANISED BY
BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH
European Commission
ELTE EÖTVÖS LORÁND UNIVERSITY
IDEAL-IST
NTP Neumann Technology Platform
HUN REN
SZTAKI
AI & AUT EXPO