



Hamburg, 16 May 2024

AI Factories

*Juan Pelegrin
DG CONNECT
European Commission*

Context

- **Generative AI** and **AI foundation models** are advancing at unprecedented pace and are set to play a pivotal role in shaping the future of technology and society.
- **AI** is a key policy area of the **EU digital strategy**. AI in combination with **HPC** can contribute to a more innovative, efficient, sustainable and competitive economy, while also improving safety, education and healthcare for citizens.



- President von der Leyen “**the supercomputing resources of the EuroHPC JU will be made available to European AI startups to train their large-scale models**, contributing to the EU’s aim of leading global advances in AI and of achieving responsible and ethical innovation”.
[2023 State of the Union address]

AI in the EU: AI Act and AI Innovation Package

The AI Act

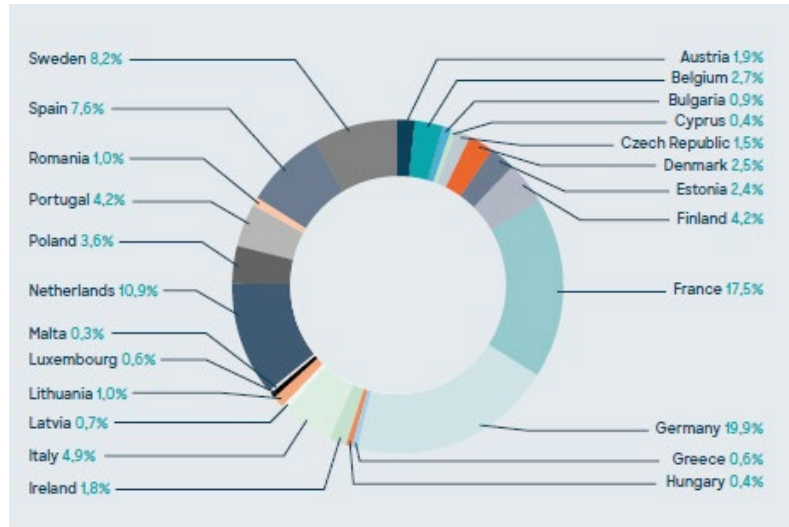
- **Product safety** and **risk-based** approach
- Protection of **health, safety** and **fundamental rights**
- A **horizontal** act
- **Coherence and complementarity** with existing legislation
- **Innovation friendly**
- Will apply to **public and private** actors, **inside and outside the EU** (as long as the AI system or general-purpose AI model is placed on the Union market), **providers and deployers**



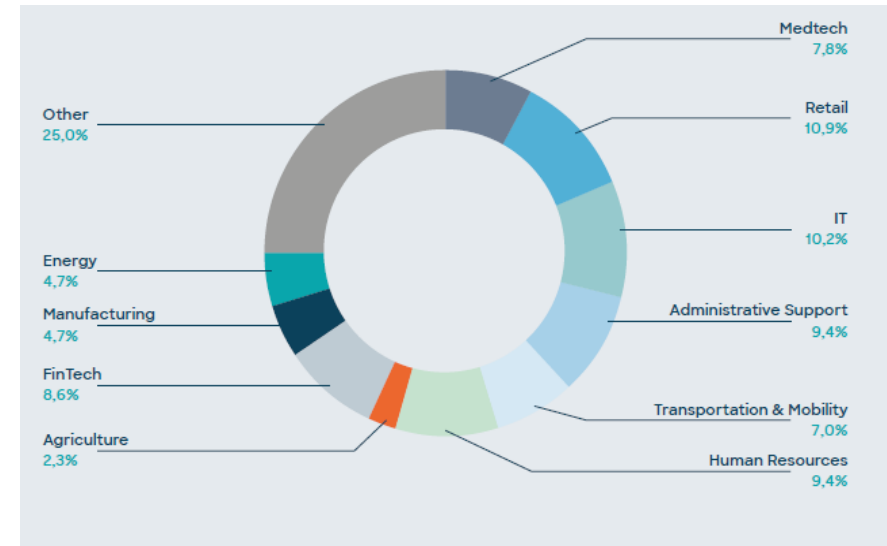
The AI Innovation Package

- **Boosting startups and innovation in trustworthy AI**
- **Two main objectives**
 - (1) AI Factories:* Making available HPC computing capacity to facilitate the development of GenAI models
 - (2) GenAI4EU:* Stimulating the development in strategic sectors of novel and innovative applications based on GenAI models and facilitating their uptake

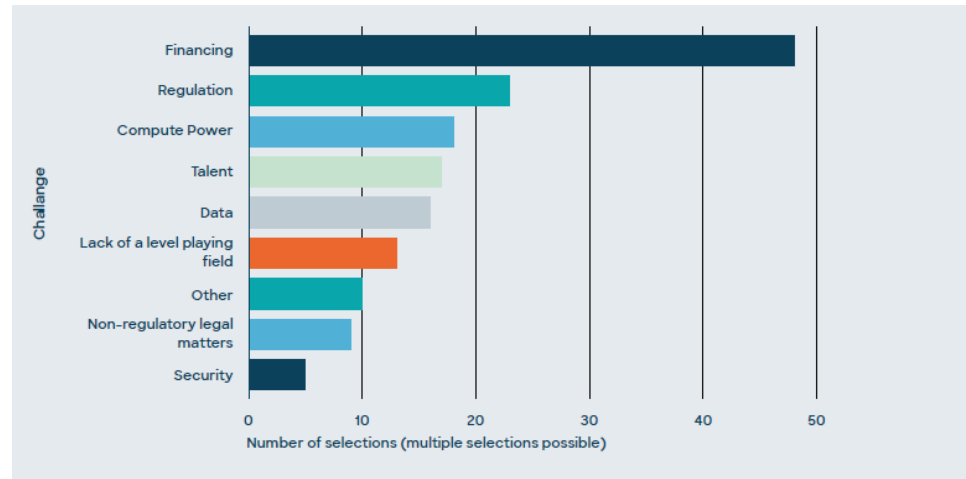
The AI Startup Landscape in the EU



Geographical distribution of the ~665 GenAI start-ups in Europe



Downstream Applications – Industry distribution

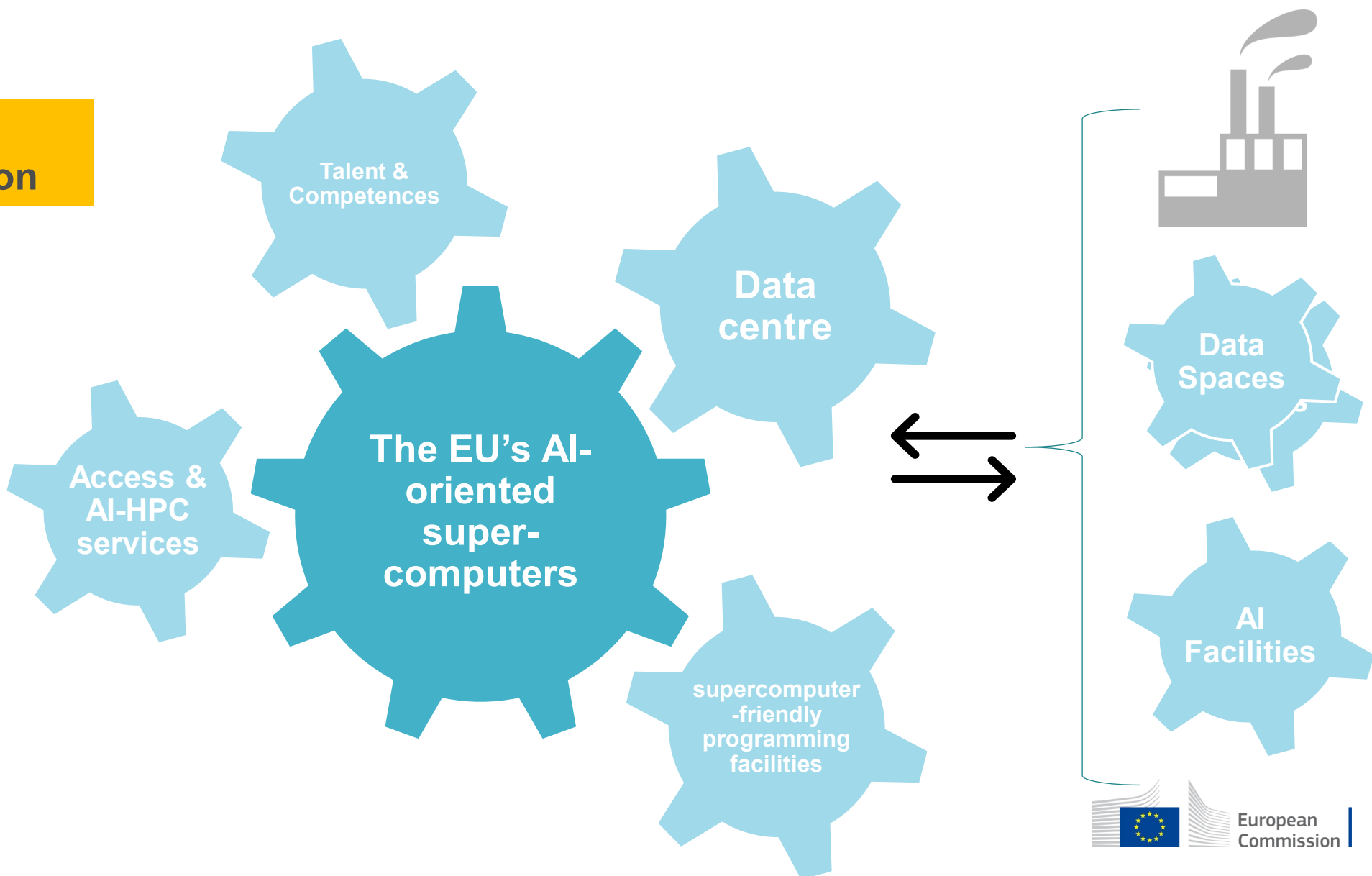


Challenges of GenAI start-ups






The AI Factories

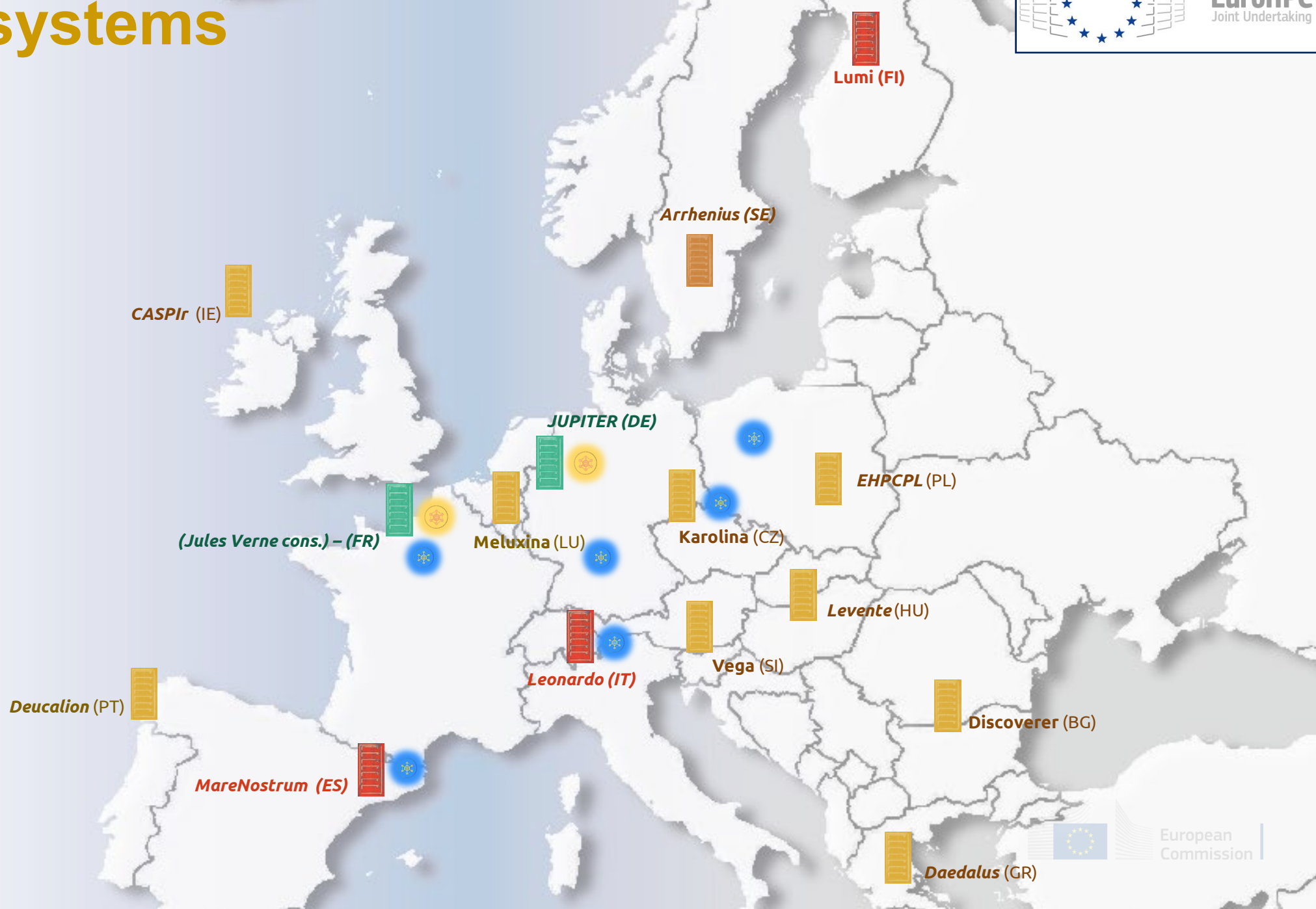
**EuroHPC
Amended Regulation**

One Stop-Shop



EuroHPC systems

-  Exascale
-  Pre-exascale
-  Petascale
-  Qcomputer
-  Qsimulator



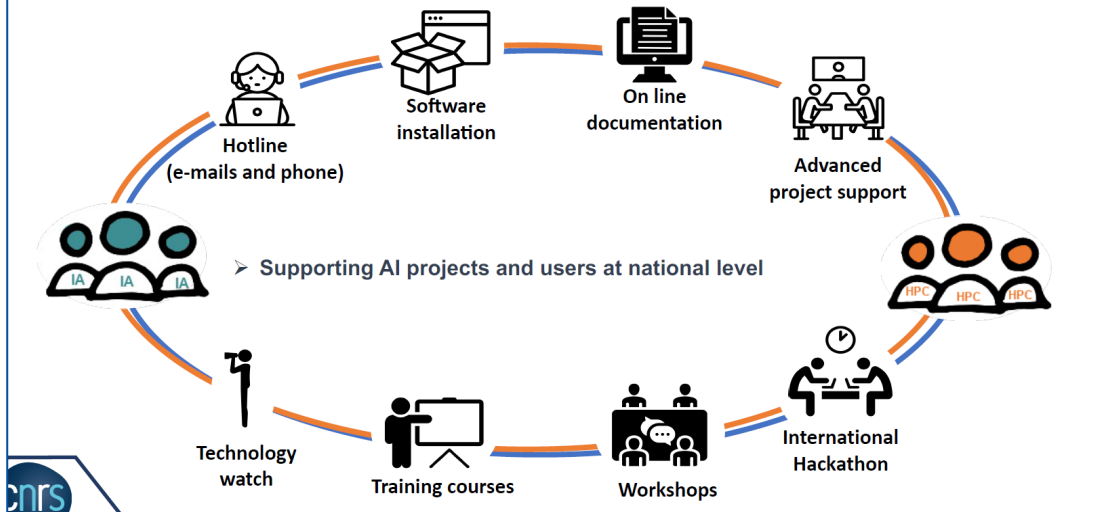
Amendments to EuroHPC Regulation (EU) 2021/1173

AI Factory - Activities

1. Acquisition & operation of AI dedicated supercomputers (co-located with a data centre)
2. Upgrading with AI existing EuroHPC
3. Provide access to SMEs and start-ups (incl. widening usage)
4. AI supercomputing service centre (algorithms, training- testing- evaluation- validation of AI models, development of large-scale AI applications, ...)
5. Supercomputer-friendly programming facilities (parallelization, usage optimization, ...)
6. Attracting & pooling talent
7. Interacting with AI-ecosystem at large & other AI initiatives

AI Factories: An example, IDRIS / Jean Zay (FR)

Missions of user support teams



- Shared disk space accessible to all users
 - Storage of voluminous data bases (in size or number of files)
 - Storage of 🤗 **huggingface_hub** models
 - More than 220 models and data bases
 - More than 1,4 Po data
 - More than 600 millions files and directories

Jean Zay computing environment



Training courses since 2021

- Practical Introduction to Deep Learning (IPDL)
 - 5 sessions, 100 people
- Optimised Deep Learning on Jean Zay
 - 5 sessions, 100 people
- FIDLE (Introduction to Deep Learning)
 - 3rd season, 20 sequences of 2 hours, 40 000 hours watched for the 2nd season
- IDRIS-NVIDIA International Hackathon for IA and HPC
 - 3 sessions, 27 projects, next session : <https://www.openhackathons.org>

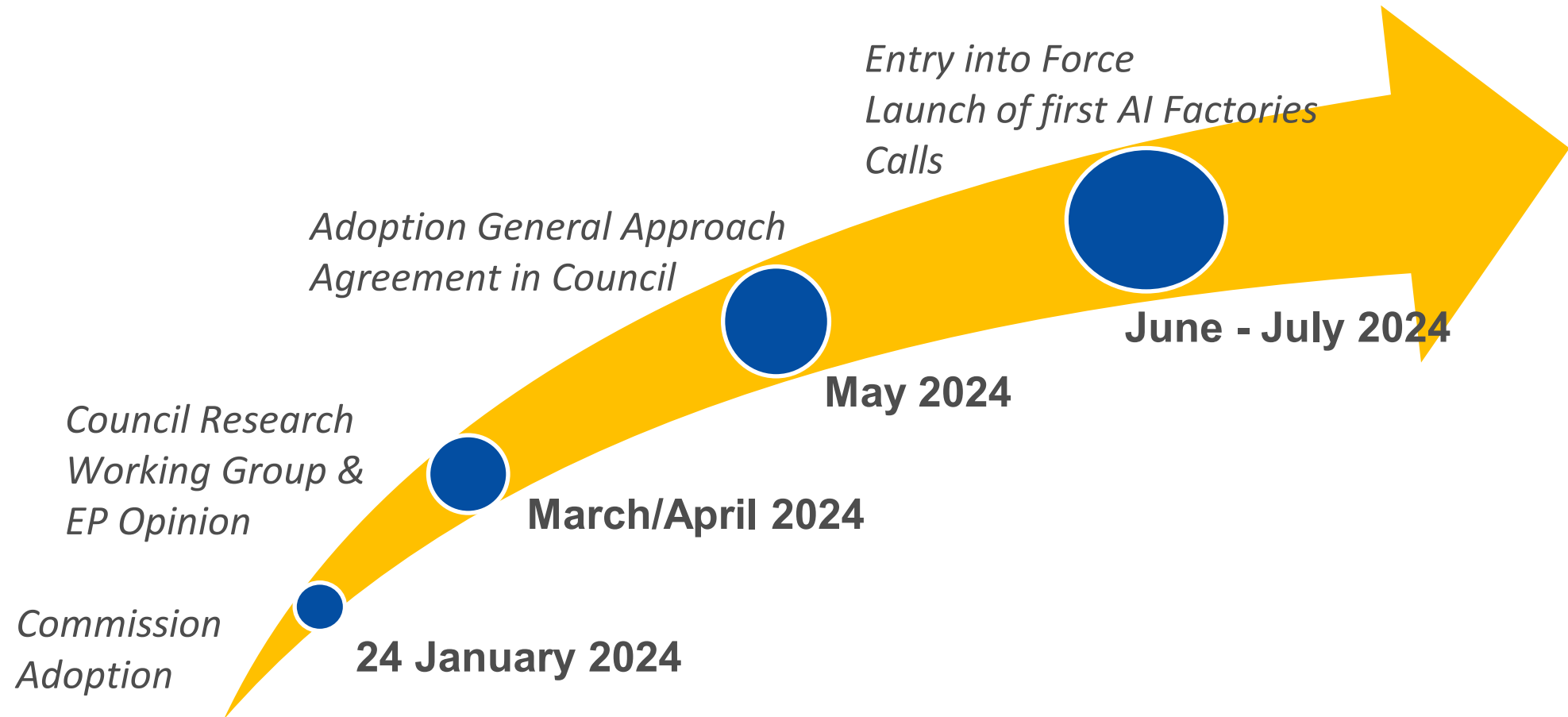


Ingredients of success for an AI Factory

- ❖ An infrastructure (HW/SW/Tools) adapted to user need
- ❖ A high-performance user support system close to the Jean Zay supercomputer



EuroHPC Amendment → Timeline



THANK YOU