PHAROS

THE GREEK AI FACTORY

Apply for HPC Access

Dr. Nikos Bakas, GRNET

EuroHPC Joint Undertaking

Co-funded by the European Union

AI Factories PHAROS GREECE

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Ψηφιακής Διακυβέρνησης και Τεχνητής Νοημοσύνης

The European High Performance Computing Joint Undertaking (EuroHPC JU)

is a joint initiative between the EU, European countries and private partners to develop a World Class Supercomputing Ecosystem.

https://eurohpc-ju.europa.eu/index_en

BDV BIG DATA VALUE ASSOCIATION

QIC European Quantum Industry Consortium

ETP 4 HPC EUROPEAN TECHNOLOGY PLATFORM FOR HIGH PERFORMANCE COMPUTING

8 operational systems, all ranking among the world's most powerful supercomputers:

1. LUMI in Finland #5
2. LEONARDO in Italy #6
3. MARENOSTRUM in Spain
4. VEGA in Slovenia
5. MELUXINA in Luxembourg
6. KAROLINA in Czechia
7. DEUCALION in Portugal
8. DISCOVERER in Bulgaria

Underway:
JUPITER in Germany
DAEDALUS in Greece

1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE
2	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel
3	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft
4	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu
5	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE
6	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, EVIDEN
7	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM

LUMI
FINLAND

LEONARDO
ITALY

MELUXINA
LUXEMBOURG

KAROLINA
CHECH REPUBLIC

DISCOVERER
BULGARIA

VEGA
SLOVENIA

DEUCALIO
PORTUGAL

MARENOSTRUM 5
SPAIN

1. **LUMI (CSC, Finland)**
   - LUMI-C: 1536 nodes, 128 cores/node, 256-1024 GB RAM/node
   - GPU: 2560 nodes, 64 cores/node, 4 GPUs, 128 GB GPU-RAM
   - Visualization: 64 nodes, 1 GPU, 48 GB GPU-RAM
   - Peak Performance: 550 petaflops
   - URL: https://www.lumi-supercomputer.eu/lumis-full-system-architecture-revealed/
2. **Leonardo (Cineca, Italy)**
   - Booster Module: 3456 nodes, 32 cores/node, 512 GB RAM/node, 4 GPUs, 64 GB GPU-RAM
   - Data Centric Module: 1536 nodes, 112 cores/node, 512 GB RAM/node
   - Peak Performance: 323.4 petaflops
   - URL: https://leonardo-supercomputer.cineca.eu/hpc-system/
3. **MareNostrum 5 (Barcelona Supercomputing Center, Spain)**
   - General Purpose Partition: 6408 nodes, 112 cores/node, 256 GB RAM/node
   - Accelerated Partition: 1120 nodes, 64 cores/node, 512 GB RAM/node, 4 GPUs, 64 GB GPU-RAM
   - Peak Performance: 314 petaflops
   - URL: https://www.bsc.es/innovation-and-services/marenostrum/marenostrum-5

4. **MeluXina (LuxProvide, Luxembourg)**
   - Cluster: 573 nodes, 128 cores/node, 512 GB RAM/node
   - Accelerator-GPU: 200 nodes, 64 cores/node, 512 GB RAM/node, 4 GPUs, 40 GB GPU-RAM
   - Large memory: 20 nodes, 128 cores/node, 4096 GB RAM/node
   - Peak Performance: 18.29 petaflops
   - URL: https://docs.lxp.lu/system/overview/
5. **Karolina (IT4I, Czech Republic)**
   - CPU: 828 nodes, 128 cores/node, 256-24000 GB RAM/node
   - GPU: 72 nodes, 8 GPUs, 40 GB GPU-RAM
   - Peak Performance: 15.69 petaflops
   - URL: https://www.it4i.cz/en/infrastructure/karolina

6. **Vega (IZUM, Slovenia)**
   - GPU partition: 60 nodes, 128 cores/node, 512 GB RAM/node, 4 GPUs, 40 GB GPU-RAM
   - CPU node Standard: 768 nodes, 128 cores/node, 256 GB RAM/node
   - CPU node Large Memory: 192 nodes, 128 cores/node, 1000 GB RAM/node
   - Peak Performance: 10.05 petaflops
   - URL: https://doc.vega.izum.si/architecture/
7. **Deucalion (Guimarães, Portugal)**
   - ARM cluster: 1632 nodes, 48 cores/node
   - X86 cluster: 500 nodes, 48+ cores/node
   - Accelerated partition: 33 nodes
   - Peak Performance: 10 petaflops
   - URL: https://macc.fccn.pt/resources#deucalion
8. **Discoverer (Sofia Tech Park, Bulgaria)**
   - CPU: 1128 nodes, 128 cores/node, 256 GB RAM/node
   - CPU-Fat: 18 nodes, 128 cores/node, 1000 GB RAM/node
   - Peak Performance: 5.94 petaflops
   - URL: https://docs.discoverer.bg/resource_overview.html

ARIS – HPC Infrastructure in Greece Compute Nodes

The ARIS infrastructure consists of a total of five computing system nodes based on Intel x86 architecture, interconnected into a single InfiniBand FDR14 network offering multiple options and processing architectures. More specifically, the infrastructure consists of:

- **Thin Nodes: 426** IBM NeXtScale **nodes**, Intel Xeon E5-2680v2, **8,520 cores**.
- **Fat Nodes: 44** Dell PowerEdge R820, **4 Intel Xeon** E5-4650v2, **512 GB memory per node**.
- **GPU Nodes: 44** Dell PowerEdge R730, **2 Intel Xeon** E5-2660v3, 64 GB memory, **2 NVIDIA K40 GPUs per node**.
- **Xeon Phi Nodes: 18** Dell PowerEdge R730, **2 Intel Xeon** E5-2660v3, 64 GB memory, **2 Xeon Phi** 7120P co-processors **per node**.
- **ML Node: 1 server**, 2 Intel E5-2698v4, 512 GB memory, **8 NVIDIA V100 GPUs.**

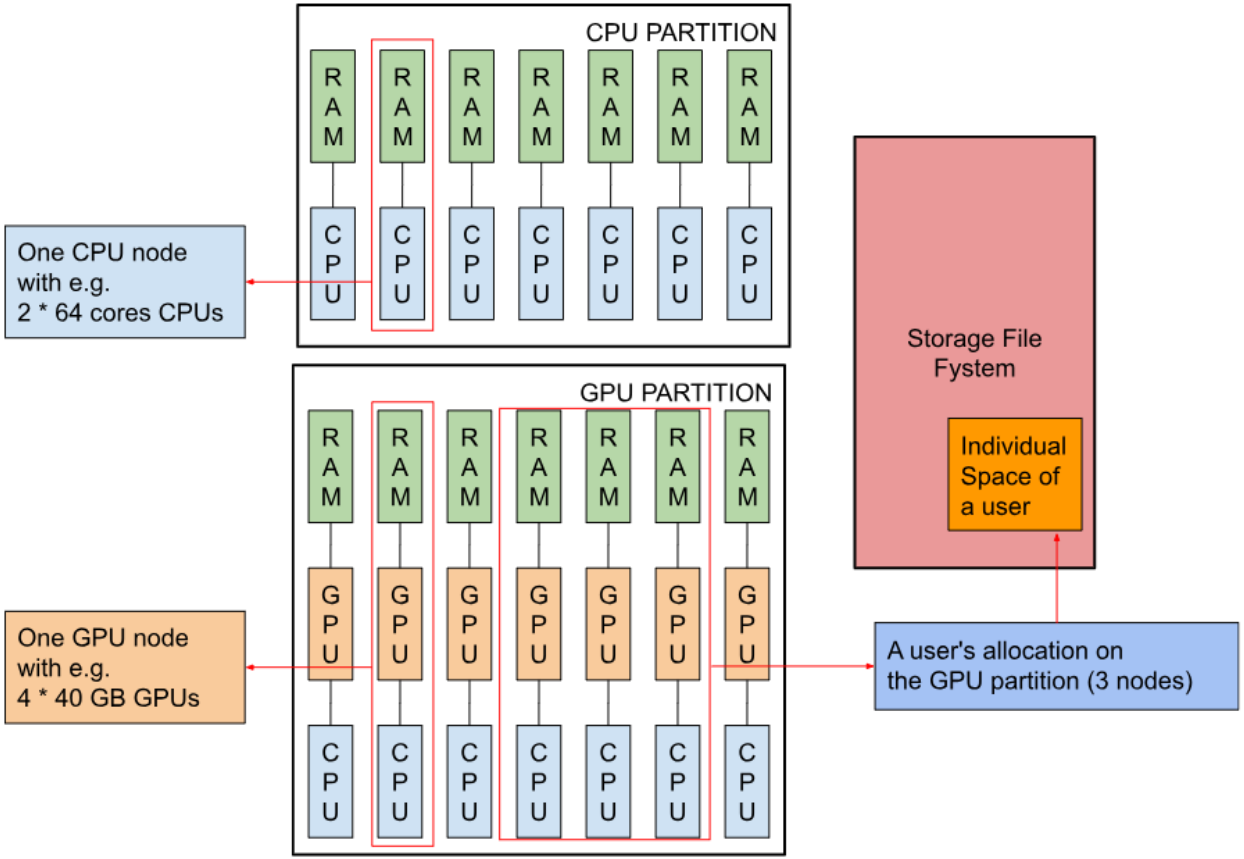
https://www.hpc.grnet.gr/

The way is open to building a EuroHPC world-class supercomputer in Greece

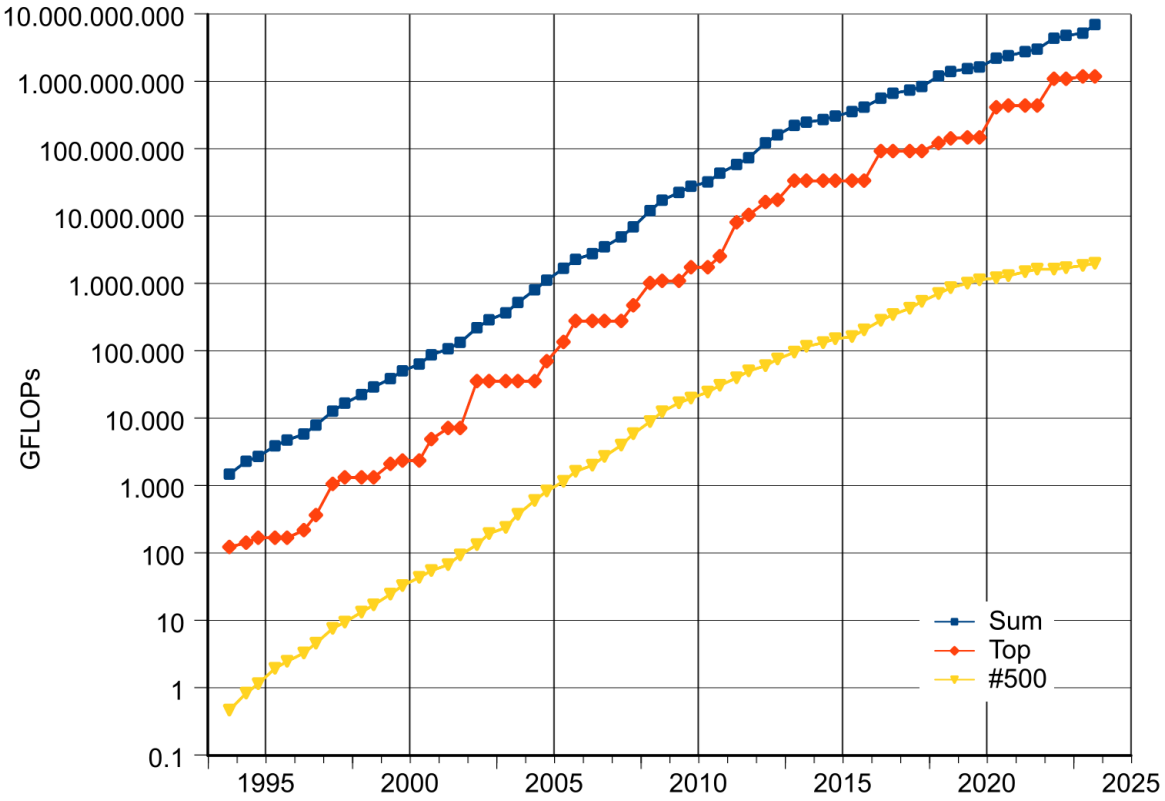
- A hosting agreement has been signed between the **EuroHPC Joint Undertaking** and the National Infrastructures for Research and Technology (**GRNET**) in **Greece**, where **DAEDALUS**, a new **EuroHPC supercomputer**, will be located.

- The DAEDALUS supercomputer, with a total power of **89 PetaFlops**, will be the most powerful computing system in Greece and one of the leading systems in Europe. https://grnet.gr/en/2025/03/26/daedalus-dc-ylopoihsh-lavrio/
- **Lavrion** Technological and Cultural Park (TCPL) https://eurohpc-ju.europa.eu/way-open-building-eurohpc-world-class-supercomputer-greece-2022-11-28_en

Example of an HPC cluster



Growth of HPC systems



The new **El Capitan** system at the **Lawrence Livermore National Laboratory** in California, U.S.A., has debuted as the most powerful system on the list with an HPL score of 1.742 EFlop/s.
https://top500.org/

https://en.Wikipedia.org/wiki/TOP500#/media/File:Supercomputers-history.svg

Supercomputing

X 200,000

0.05 km/h

1 km/h

100 km/h

1,000 km/h

10,000 km/h

1,000,000,000 km/h

1 TFlop

X 1,000,000

1 EFlop

Operations	Name	Abbreviation
1	FLOPS	FLOPS
10^3	Kilo FLOPS	KFLOPS
10^6	Mega FLOPS	MFLOPS
10^9	Giga FLOPS	GFLOPS
10^{12}	Tera FLOPS	TFLOPS
10^{15}	Peta FLOPS	PFLOPS
10^{18}	Exa FLOPS	EFLOPS

Large Language Models on HPC

Estimated GPU Hours for Training:

1.Small LLM (~8B):

- ~1.3M GPU hours (LLAMA 3 8B).

2.Medium LLM (~70B):

- ~6.4M-7.0M GPU hours (LLAMA 3/3.1 70B).

3.Large LLM (~405B):

- ~30.84M GPU hours (LLAMA 3.1 405B).
- Falcon 180B (slightly smaller): ~7M GPU hours.

Estimated GPU Requirements for Inference:

- **Small LLM (~8B):** ~80GB GPU RAM (LLAMA 3.1 8B).
- **Medium LLM (~70B):** ~320GB GPU RAM (GPTQ/int4 on Falcon).

Large LLM (~405B): ~800GB GPU RAM (FP8 on LLAMA 3.1).

https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct
https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct
https://huggingface.co/meta-llama/Llama-3.1-405B-Instruct
https://docs.lxp.lu/howto/llama3-vllm/
https://huggingface.co/blog/falcon-180b

AI Factories Access Calls - https://www.eurohpc-ju.europa.eu/ai-factories/ai-factories-access-calls_en

Playground Access to AI factories

Playground mode is intended for SMEs, startups and entry level users. It offers rapid, FIFO access to test technologies, including access within 2 working days and onboarding services for new users.

Fast Lane Access to AI factories

Fast Lane access is intended for SMEs and startups performing AI activities that require medium size allocations.

EuroHPC JU Call for Proposals for AI for Science and Collaborative EU Projects

This access mode covers all types of scientific users (whether funded or not by national or European programmes), users from public sector, as well as industrial users participating in R&I projects funded by EU Programmes such Horizon Europe or the Digital Europe Programme.

Large Scale Access to AI factories

Large Scale access mode is for AI models and applications requiring more than 50,000 GPU hours for a maximum of one year.

2025 Cut off dates for EuroHPC Access Calls

BENCHMARK ACCESS:

- 1st day of each month

DEVELOPMENT ACCESS:

- 1st day of each month

AI FOR SCIENCE AND COLLABORATIVE PROJECTS

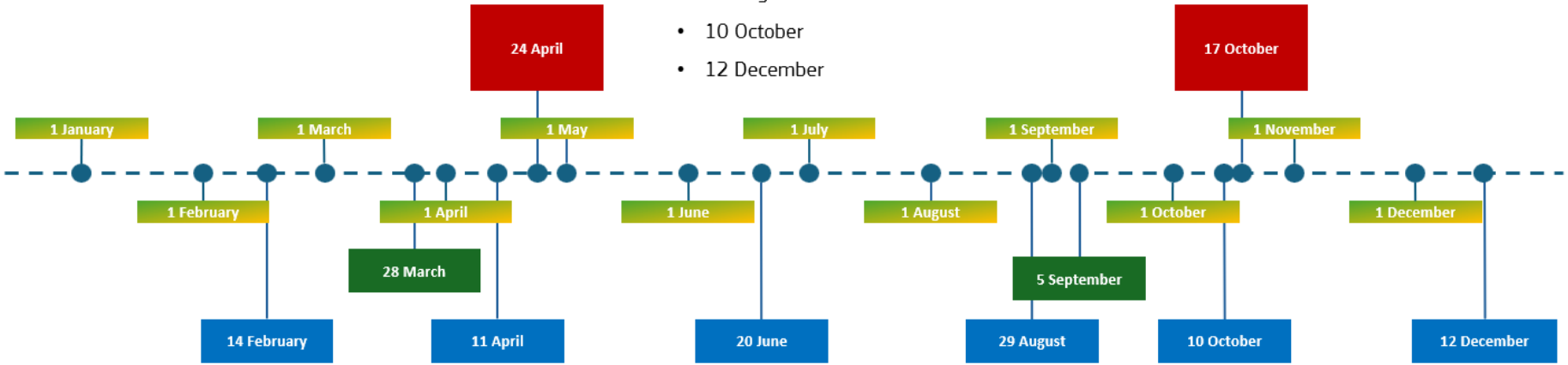
- 14 February
- 11 April
- 20 June
- 29 August
- 10 October
- 12 December

REGULAR ACCESS:

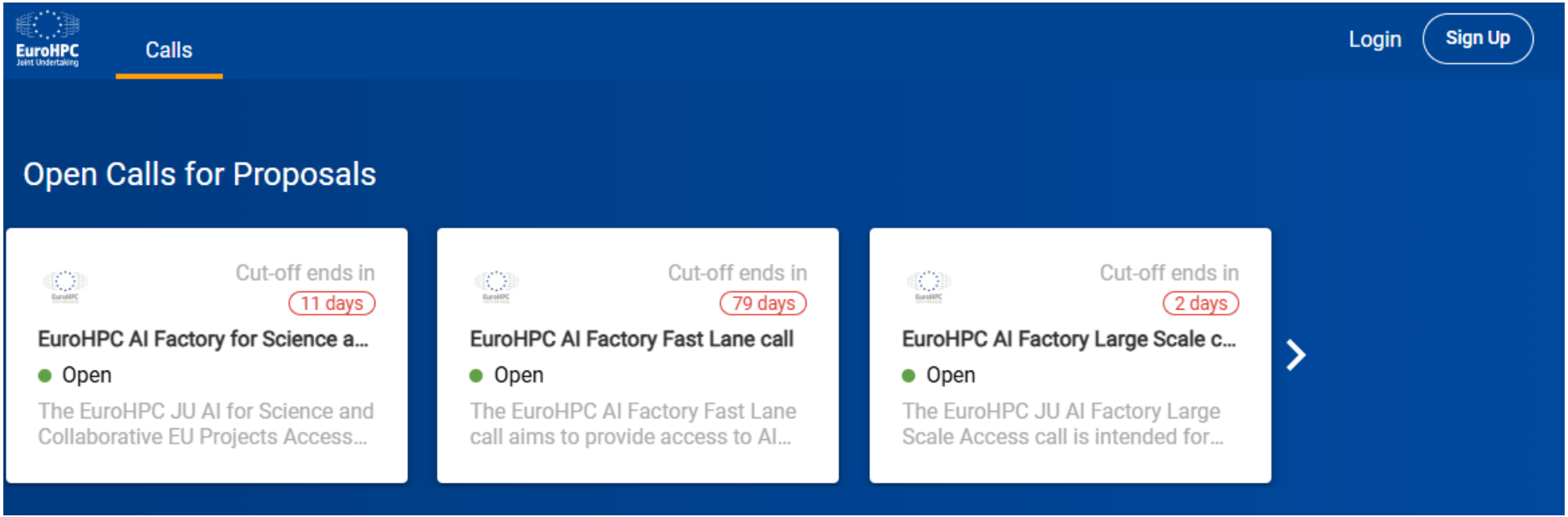
- 28 March
- 5 September

EXTREME SCALE ACCESS:

- 24 April
- 17 October



https://eurohpc-ju.europa.eu/access-our-supercomputers/access-policy-and-faq_en



https://access.eurohpc-ju.europa.eu/

Frequently Asked Questions (FAQ)

https://eurohpc-ju.europa.eu/access-our-supercomputers/access-policy-and-faq_en

- **How can I gain access to computation time on EuroHPC machines?**
  - You will need to **apply** to one of the open **access calls** that **EuroHPC** provides. The list of available calls can be found here.
- **Which organisations are eligible for access to EuroHPC machines?**
  - Any European organisation is eligible for access to perform Open Science research (the **results** of the work are made available for **open access**). This includes **public and private** academic and research institutions, public sector organisations, industrial enterprises and SMEs
- **What is the cost?**
  - Currently access is **free of charge**.
- **What are the participation conditions?**
  - Participation conditions depend on the specific access call that a research group has applied. In general users of EuroHPC systems commit to: **acknowledge** the use of the **resources** in their related publications, contribute to **dissemination** events, produce and submit a **report** after completion of a resource allocation. More information on participation conditions can be found in the call's Documents section.

Subscribe!

https://www.pharos-aifactory.eu/subscribe-form/

https://www.linkedin.com/company/pharosaifactory/

https://www.youtube.com/@PharosAIFactory

PHAROS

THE GREEK AI FACTORY

Pharos Training Series

Dr. Nikos Bakas, GRNET

EuroHPC Joint Undertaking

Co-funded by the European Union

AI Factories PHAROS GREECE

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Ψηφιακής Διακυβέρνησης και Τεχνητής Νοημοσύνης