



HPC Training Series

Course 3

Large Language Models (LLMs) on High Performance Computing (HPC) Systems

| PRESENTATION LANGUAGE: GREEK |

APRIL 24, 2024 | 10:00 EET | ONLINE



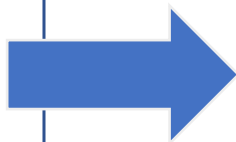
Ilias Hatzakis, GRNET S.A.



The Greek Competence Center for High Performance Computing and Artificial Intelligence

Enhancing innovation capacity in Business, Industry and Science
by utilizing advanced High Performance Computing services

SUPPORT
ACCESS
TRAINING
CONSULTING
ASSESSMENT



RESEARCH & ACADEMY
SMEs & INDUSTRY
PUBLIC ADMINISTRATION

Life sciences/ Healthcare
Biotechnologies
Materials & Batteries
Engineering & Manufacturing
Climate change & Meteorology
Cybersecurity
Robotics

AI
Machine Learning
LLMs
Simulations
Digital Twins
Large Datasets

About EuroCC@Greece



A National Competence Center is the reference and single point of contact and coordination on a national level for HPC. Its missions are to analyse, implement and coordinate all necessary activities and offer services to end users to cover their needs: from access to resources and technological consultancy to the provision of training courses for academia, public administrations and industry.

The aim is to bring together the necessary expertise to set up a cross-European network of NCCs in HPC-related topics with 31 participating members and associated states and to provide a broad service portfolio tailored to the respective national needs of academia, public administrations and industry. Each NCC has a presentation page with their skills and contact information.

EuroCC@Greece is one of the 33 HPC Competence Centres, built in the framework of the European High Performance Computing Joint Undertaking (EuroHPC JU).

The overall objective of the Greek National Competence Center is to enable the efficient uptake of HPC technologies

with the 3-fold goal to:

- i) advance competitiveness in research
- ii) improve effectiveness of government services and
- iii) promote innovation in industry.

In order to achieve this goal, the NCC will address the issues of training and skills development, technology transfer, collaboration with Industry, competence mapping and awareness raising, in the fields of High Performance Computing, High Performance Data Analytics, Artificial Intelligence and Big Data.

Consortium

The Greek National Competence Center “EuroCC@Greece”, is run by a consortium of 5 institutions, namely

- GRNET – National Infrastructures for Research and Technology (**coordinator**),
- National Center for Scientific Research “Demokritos”,
- Foundation for Research and Technology – Hellas (FORTH),
- Institute of Communication and Computer Systems of NTUA
- Aristotle University of Thessaloniki.

The project has received funding from the European High-Performance computing Joint Undertaking (JU) under grant agreement No 951732 and the Greek Secretariat for Research and Technology.



EuroHPC
Joint Undertaking

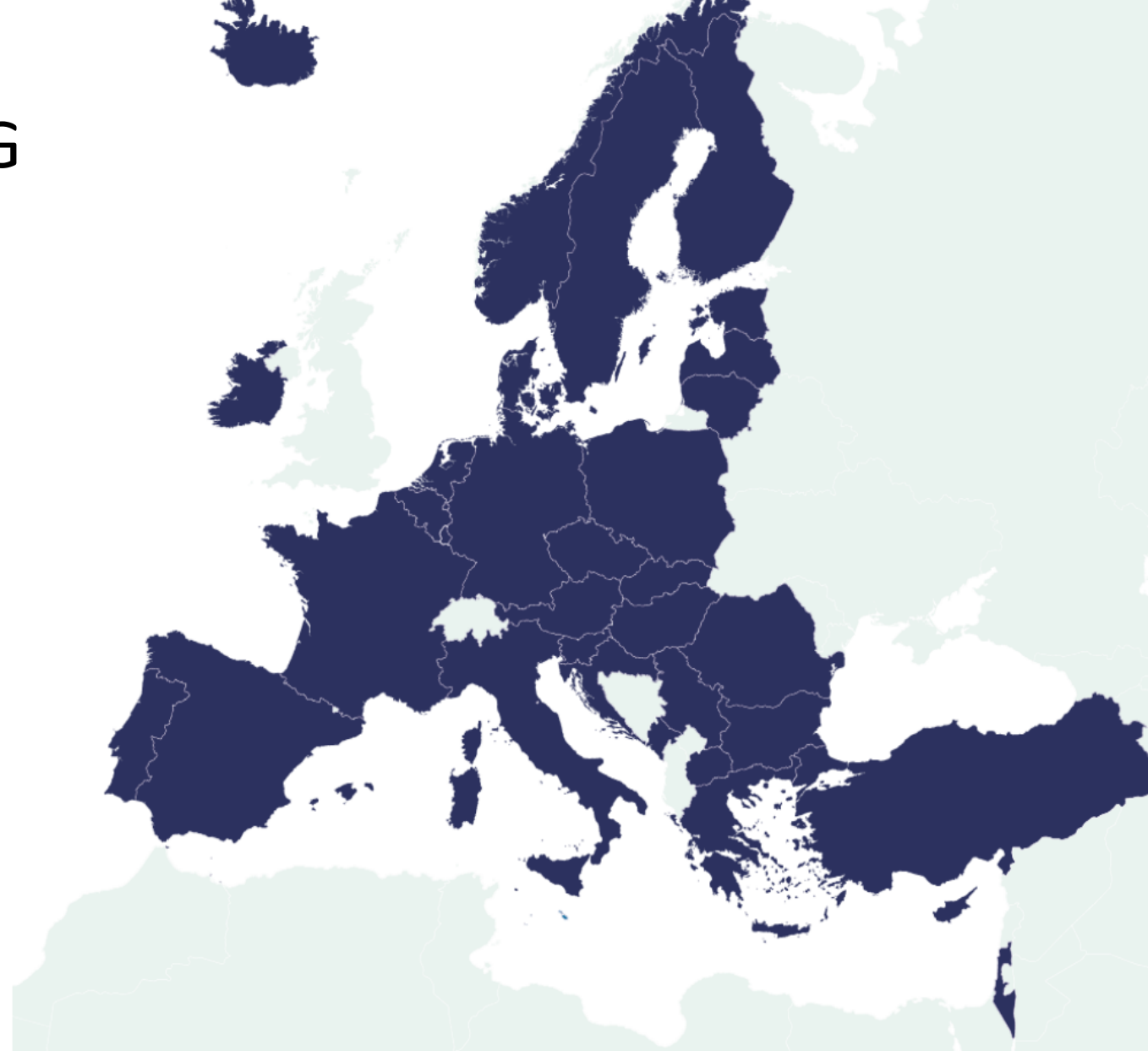


EUROHPC JOINT UNDERTAKING

- 34 participating countries
- The European Union (represented by the European Commission)
- 3 private partners

Each of our members is represented in the EuroHPC JU's Governing Board

The Governing Board also takes advice from the EuroHPC Industrial and Scientific Advisory Board (INFRAG & RIAG)





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

- LUMI in Finland #5
- LEONARDO in Italy #6
- MARENOSTRUM in Spain
- VEGA in Slovenia
- MELUXINA in Luxembourg
- KAROLINA in Czechia
- DEUCALION in Portugal
- 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



HPC ecosystem

- development and uptake of demand-oriented and user-driven innovative and competitive supercomputing system
- widen the use of that supercomputing infrastructure to a large number of public and private users and support
- development of a wide range of applications optimized for these systems;
- development of key HPC skills for European science and industry
- based on a supply chain that will ensure components, technologies and knowledge limiting the risk of disruptions



AMENDMENTS TO EUROHPC REGULATION

- **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.



- In her 2023 State of the Union address, President von der Leyen announced that **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.

Amendments to EuroHPC Regulation (EU) 2021/1173

2. AI Factory - Activities

1. Acquisition & operation of AI dedicated supercomputers (co-located with 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

HPC & AI skills development Training series



- HPC for beginners
- Εισαγωγή σε Open MP/ MPI/ CUDA
- Εισαγωγή στην Python για HPC
- Εισαγωγή στην επιστήμη δεδομένων: Structured data, Time series και TEXT
- Γενικές έννοιες παράλληλου προγραμματισμού
- Εισαγωγή σε Linux, Cluster computing και scripting
- Εφαρμογές γραμμικής άλγεβρας
- HPC with C# and Visual Studio
- Machine learning & neural networks
- Εξοικείωση με domain specific αλγόριθμους (GROMACs, OpenFOAM, PYTORCH, TENSORFLOW)
- <https://docs.google.com/forms/d/e/1FAIpQLSdHpJyAbLuWwpelBy812mEirEOGzYnV5JFMnSQkCu4atMnJA/viewform>

3^ο ΕΚΠΑΙΔΕΥΤΙΚΟ ΣΕΜΙΝΑΡΙΟ

EURO
Greece

it.auth

HPC Training Series

Course 3

Large Language Models (LLMs) on High Performance Computing (HPC) Systems

| PRESENTATION LANGUAGE: GREEK |

APRIL 24, 2024 | 10:00 EET | ONLINE

Thank you!



EuroHPC
Joint Undertaking

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 951732. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, United Kingdom, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Switzerland, Turkey, Republic of North Macedonia, Iceland, Montenegro