MINERVA

European Support Centre for Scalable AI Research and Deployment

Webinar #2 – AI for SMEs – 20 Jan 2026

About the Moderator

Luigi Russi

MINERVA WP6 Leader

HPC Technology Specialist

CINECA

About Us

Cineca is a non profit Consortium, made up of 119 members: 2 Ministries, 70 Italian Universities and 48 Italian National Institutions and agencies.

The High Performance Computing (HPC) department of CINECA is the largest computing centre in Italy and one of the largest in Europe.

The mission of the HPC Department is to accelerate scientific discovery by providing high-performance computing resources, data management and storage systems and tools and HPC services and expertise at large, aiming to develop and promote technical and scientific services related to high-performance computing for the Italian and European research community.

CINECA enables world-class scientific research by operating and supporting leading-edge supercomputing technologies and managing a state-of-the-art and effective environment for the different scientific communities. The HPC staff offers support and consultancy in HPC tools and techniques in several scientific domains, such as physics, particle physics, material sciences, chemistry and bioinformatics.

Webinar #2 - Agenda

- Welcome and Project Introduction *Claudio Arlandini (CINECA)*
- SMEs & AI *Beatrice Del Frate (Domyn)*
- MINERVA Services *Kamel Guerda (IDRIS – CNRS)*
- Success Stories *Alexandros Paliouras (BSC)*
- Contacts & Discussion *Claudio Arlandini (CINECA)*

About MINERVA

Luigi Russi

Objectives

MINERVA is a distributed, European-wide HPC-enabled AI application support service

It brings together expertise from EuroHPC Hosting Entities and partners representing major European stakeholders in AI.

MINERVA acts as a central hub for cutting-edge European competences in large-scale ML/AI research and development.

It started in January 2025, and the project's duration is 36 months.

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101182737. The JU receives support from the Digital Europe Programme.

Partners


MINERVA


INES
Centre Informatique National
de l'Enseignement Supérieur


CINECA


UNIVERSITAS STUDIORUM MUTINENSIS ET REGIENSIS
1175
UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA


cnrs


IDRIS


AMD
SILO AI


cea


JÜLICH
Forschungszentrum


Tübingen AI Center


Domyn


BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Who can access MINERVA support

MINERVA services are open to all actors in the European AI ecosystem – regardless of technical background or experience with HPC. If you're working with AI and face technical, scaling, or regulatory challenges, MINERVA is here to help.

We welcome requests from:

- **Academic researchers** (PhD students, postdocs, principal investigators)
- **Public sector teams** applying AI in areas like health, climate, education, or policy
- 👉 **Start-ups and SMEs*** developing AI-driven products or services
- **Larger companies*** with advanced or specialized AI use cases
- **AI engineers and developers** aiming to scale models or optimize performance
- **Project consortia** preparing proposals for national or European funding calls

📌 *No HPC access? We can guide you through the process of applying for resources, choosing the right infrastructure, and getting started with your first project.*

* subject to state-aid regulations

SMEs & AI

Beatrice Del Frate

About the Presenter

Beatrice Del Frate

MINERVA T3.4 Leader

AI Product Manager

Domyn

Driving innovation through AI expertise

Domyn creates sovereign, responsible AI that leaders can adopt and apply with full autonomy and control.

- 150+ team members
- 25 nationalities represented
- 15 languages spoken
- HQs in Milan and New York

Relevance of small businesses

Small businesses, and in particular Small-Medium Enterprises (< 250 employees), are of particular interest for MINERVA because they pose peculiar challenges comparing to other targets of the Project.

Huge Market Relevance...
SMEs are a key driver of the European economy: *"Micro & small businesses make up 99% of enterprises in the EU"*.

...but AI Adoption Struggle
At the same time, SMEs are the type of companies that struggle the most to adopt AI systemically: *"SME AI adoption remains relatively low compared to other digital technologies and to larger firms."*

They are, however, not alone: the *"pilot-to-production chasm"* states that, across all types of organizations, *"only 5% of custom enterprise AI tools reach production"*, remaining stuck at the Proof-of-Concept stage!

Sources
- European Commission
- Eurostat (2024)
- OECD (December 2025)
- MIT, "The GenAI Divide STATE OF AI IN BUSINESS" (2025)

The heart of the problem

At MINERVA we know that the problem is not lack of interest and believe that it instead consists of the presence of concrete obstacles that act as ***entry barriers to successfully bring AI systems into production***.

For example:

- 👥 lack of internal competences; difficulty to find specialized profiles
- 🛠️ lack of usage / application clarities
- 💻 lack of computational resources
- 💰 concern with high costs of training
- 🔐 concern with security & privacy and lack of trust with most AI providers given sensitiveness of data, especially in regulated industries (manufacturing, financial services, health, etc.)
- 📚 concern with General-Purpose models; need for tailored solutions in terms of domain knowledge

Sounds familiar...?

Time for discussion #1

MINERVA

Concrete scenarios

- 🛠️ Manufactoring company has vast amount of data, but cannot access and understand them as they are scattered from different sources / databases
- 🔐 HR company trying to optimize candidates screening, but fearing GDPR & AI Act obligations
- 📚 Fintech wants to automate risk analysis and loans process, but doesn't trust a model that is not "finance fluent"
- 💰 AI company experiments to train a LLM model without parallelism & optimization strategies, and costs skyrocket

🎤 **Let us know your point of view!**

- How are you trying to apply AI in your company?
- Which obstacles or concerns do you struggle with the most?
- Would you want to share a concrete scenario with us?

MINERVA Services

Kamel Guerda
Beatrice Del Frate
Alexandros Paliouras

About the Presenter

Kamel Guerda

MINERVA WP2 Leader

AI & HPC Research Engineer

cnrs IDRIS

Scaling AI with France's national HPC expertise

IDRIS provides production-grade HPC services and expertise to accelerate AI research and engineering at scale.

- National HPC center under CNRS
- Host of **Jean Zay** (since 2019)
- AI & HPC support: training, project support, collaborations

Why MINERVA services matters for your AI

At MINERVA we have the ambition to demolish some of these entry barriers by offering *guided support to run AI applications on European supercomputers*, making AI more accessible, efficient, and tailored to each company.

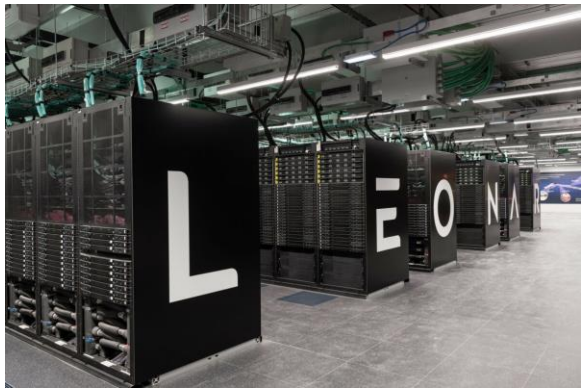
What we offer:

- 💻 Guidance on accessing European Computational Resources
- 💰 Speed & Cost Efficiency
- 🔒 Compliance & Safety By Design
- 📚 Tailored AI Models / LLMs
- 🤵 Multi-Level Support

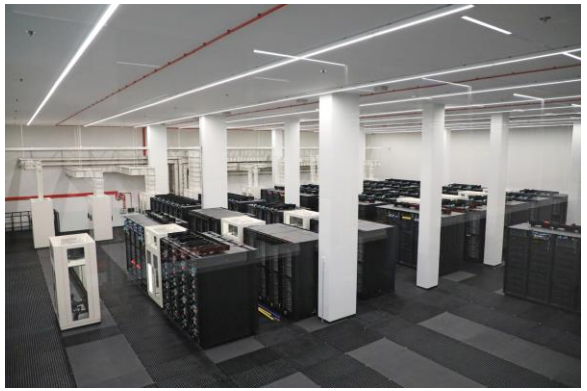
Let's deep dive...

Why MINERVA matters for your AI

- 💻 **Access to European Computational Resources**
  - We support AI teams in accessing and effectively leveraging European ***High-Performance Computing (HPC)*** infrastructures



Leonardo Supercomputer,
CINECA (Italy)



MareNostrum5 Supercomputer,
BSC (Spain)

Jean Zay Supercomputer,
IDRIS (France)

We help you master AI libraries beyond basic porting level..

Why MINERVA matters for your AI

- 🧑‍🏫 **Multi-Level Support**
  - Assistance: Quick resolution of specific issues
  - Consulting: Strategic guidance for planning and theoretical questions
  - Training: Accelerate workflows through hands-on engineering support (e.g. Workshops, AI Winter School2026)
  - Embedded Support: Long-term collaboration on advanced AI projects
- 💰 **Costs & Speed Efficiency**
  - Achieve results faster and cheaper: no need to reinvent the wheel
  - We can help with funding proposals
- 🔐 **Compliance & Safety By Design**
  - Address challenges stemming from EU regulations and best practices in ethical and responsible AI
- 📚 **Tailored AI Models / LLMs**
  - We provide guidance in the customization of large open foundation models, from pre-training to fine-tuning and specialization methods

MINERVA: objectives and delivery model

What we do

- Offer a rich service portfolio covering several levels and types of support.
- Interact with AI communities to identify needs and update the MINERVA service portfolio.
- Increase competences of AI communities on AI n HPC.
- Ensure models are developed according to ethical and responsible AI regulations.

How we do it

- **Support:** Operating an EU-wide support centre.
- **Knowledge transfer**: Publishing best practice guides and guidelines.
- **Benchmarking**: Evaluating model performance on supercomputers.
- **Data Access**: Providing information on access to public datasets.
- **Training Programs**: Providing "AI on HPC" training program.
- **Community Hub**: Supporting large-scale open-source ML/AI research and development on HPC.

Time for discussion #2

🎤 **Let us know your point of view!**

- Which kind of support would you need the most?
- Have you tried to solve your problems before and how?
- Do you see potential for a collaboration with MINERVA? How do you imagine it?

SUPPORT LEVELS AND TYPES

MINERVA

L1
Porting AI Applications & Workflows to HPC

L2
Optimization & Scaling of AI Libraries on HPC

L3
Pre-training of Large-Scale Foundation Models

L4
Specialization & Fine-Tuning of Foundation Models

L5
Ethical & Responsible AI Support

TYPE 1
Assistance

TYPE 2
Consulting

TYPE 3
Workshop

TYPE 4
Embedded Support

T1 – Assistance: First-Line User Support

Reactive, short interactions to unblock users on concrete issues (jobs, configs, environments, resources).

- **Example with L1 – Porting**
  - ⚠️ Problem: Doctoral researcher gets Slurm errors when launching a training job on an HPC cluster.
  - 💡 Support: MINERVA reviews the job script and logs to identify misconfigured directives.
  - 💡 Support: We provide a corrected script and show how to submit and monitor future jobs.

- **Example with L2 – Scaling**
  - ⚠️ Problem: SME engineer observes unexpectedly long training times on a multi-GPU setup.
  - 💡 Support: MINERVA suggests profiling tools to locate bottlenecks in the training workflow.
  - 💡 Support: We recommend input-pipeline changes to fix data loading and improve GPU utilization.

- **Example with L3 – Specialization**
  - ⚠️ Problem: Team wants to fine-tune a language model but is unsure which model and datasets to pick.
  - 💡 Support: MINERVA leverages ongoing benchmarks and partners' expertise to suggest suitable open models.
  - 💡 Support: We highlight relevant domain datasets and share references to kick-start the fine-tuning effort.

T2 – Consulting: Strategic Guidance Sessions

Structured, planned sessions to provide expert guidance on architecture, scaling, and regulatory questions.

- **Example with L2 – Scaling**
  ⚠️ Problem: Remote-sensing team wants to scale a high-resolution segmentation model to larger datasets and higher input resolution on HPC.
  💡 Support: MINERVA performs targeted technology watch on distributed training and I/O bottlenecks for similar use cases.
  💡 Support: We give tailored advice on dataset handling, storage layout, and data/model/pipeline parallelism, plus suitable EuroHPC/national resources.

- **Example with L5 – Ethical & Regulatory Compliance**
  ⚠️ Problem: Start-up building clinical decision-support AI needs clarity on how the EU AI Act applies to its use case.
  💡 Support: MINERVA reviews documentation and identifies likely risk classification and applicable obligations.
  💡 Support: We walk through key requirements, share best-practice examples, and provide a checklist for documentation and future audits.

T3 – Workshop: Hands-On Engineering Sessions

Intensive 1–2 day hands-on sessions to work directly on users' code and workflows on HPC systems.

- **Example with L2 – Optimization**
  - ⚠️ Problem: Research lab on satellite object detection faces inefficient training on a GPU cluster.
  - 💡 Support: MINERVA reviews training scripts and profiling traces ahead of the workshop.
  - 💡 Support: Together we apply mixed precision, async data loading, and improved GPU scheduling to boost throughput.

- **Example with L3 – Fine-Tuning**
  - ⚠️ Problem: Private SME with limited AI experience wants to fine-tune a language model on its internal documentation.
  - 💡 Support: MINERVA reviews the existing data, infrastructure, and target use cases before the workshop.
  - 💡 Support: In the session we design the fine-tuning pipeline, apply PEFT (e.g. LoRA), and define evaluation on real internal queries.

T4 – Embedded Support: Medium to Long-Term Collaboration

Embedded collaboration over several months, with a MINERVA engineer working inside the team on development, optimisation, and research around complex AI workflows.

- Example with L1 – Workflow Adaptation
  - ⚠️ Problem: An SME R&D team needs to migrate a CPU-based AI workflow to a GPU-accelerated HPC system to reduce training time and scale experiments.
  - 💡 Support: A MINERVA engineer is embedded for three months to review the codebase and refactor the workflow for GPU execution.
  - 💡 Support: Together we build a reproducible pipeline with experiment versioning, leading to a stable, documented HPC setup.
- Example with L3 – Pre-training
  - ⚠️ Problem: The legal services of a large company wants to fine-tune an LLM for contract review and Q&A over internal policies, under strict confidentiality constraints.
  - 💡 Support: A MINERVA engineer is embedded to prepare datasets, run fine-tuning on EuroHPC resources, and set up evaluation.
  - 💡 Support: We integrate logging/monitoring, ensure compliance with EU data-handling rules, and transfer best practices from MINERVA benchmarks.

A typical user journey

How a user can move from first access to HPC to advanced collaboration.

1. First contact – Getting onto HPC (T1 Assistance)
   A bioinformatics postdoc contacts MINERVA to access EuroHPC for the first time. Support helps with resource application, environment setup, and first job submissions.

2. Second contact – Becoming efficient (Docs, training, T3 Workshop)
   They use MINERVA guides and training to work more autonomously. When performance issues appear, a T3 Workshop is organised to profile the code and optimise GPU and data loading.

3. Third contact – Planning and scaling (T2 Consulting)
   As the project grows, they request T2 Consulting to discuss scaling to larger datasets, storage and reproducibility, and relevant EuroHPC calls, leaving with a concise roadmap.

Best Practice Guidelines – 2 collections

Soon available on MINERVA website

Leading:
Domyn AI Research Team

A project that aims at providing the European community of AI developers and researchers / AI practitioners with comprehensive, reliable, and updated online resources about how to develop AI models on European HPC infrastructures.

"Optimizing AI Models on European Infrastructure"

- 📕 *"Infrastructure Foundations for Scalable LLM Training"*
- 📗 *"Efficient and Reliable LLM Pretraining"*
- 📘 *"Post-Training: Alignment and User-Centered Objectives"*
- 📙 *"Explainable AI"*

Status: Writing
Publication: June 2026

"Dataset Management for European AI Ecosystems"

- 📕 *"Dataset Sourcing: Where and How to Collect Data"*
- 📗 *"Cleaning, Preprocessing, and Curation"*
- 📘 *"Quality, Ethics, and Compliance"*
- 📙 *"Publishing, Reuse, and Maintenance"*

Status: Writing
Publication: December 2027

About the Presenter

Alexandros Paliouras

MINERVA WP3 Co-Lead

AI Engineer

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS) is the national supercomputing centre in Spain. We specialize in high performance computing (HPC) and manage MareNostrum, one of the most powerful supercomputers in Europe.

BSC is at the service of the international scientific community and of industry that requires HPC resources. Our multidisciplinary research team and our computational facilities –including MareNostrum– make BSC an international centre of excellence in e-Science.

Benchmarking

Motivation & Need for Benchmarking

- As AI models grow in size and complexity, selecting the right supercomputing system and configuration/frameworks becomes critical for performance, cost and energy efficiency.
- Without standardized benchmarks, comparisons across systems are **subjective, inconsistent and non-reproducible.**
- Goal: Create our custom benchmarks to provide **transparent**, **data-driven insights** to help researchers and institutions choose the best hardware and software setup for their AI workloads.

Inference Benchmarks:

- **Focus:** Serving phase of pre-trained models
- **Key metrics:** Latency, Throughput, Energy
- **Frameworks across EuroHPC systems:**
  - vLLM, deepspeed-mii, SGLang

Training Benchmarks:

- **Focus:** Learning/Fine-Tuning phase
- **Key metrics:** Throughput, Convergence Time, Energy consumption, Scaling Efficiency
- **Frameworks across EuroHPC systems:**
  - Torchrun, accelerate, deepspeed

Benchmarking

🔬 **Benchmark Expansion**

- **Integrate standardized benchmarks**:
  - **MilaBench** – Advanced LLM Training benchmarking suite.
  - **MLPerf** – Industry-standard ML performance benchmarks for Inference/Training.
  - Others

- **More diverse workloads**:
  - Reinforcement Learning.
  - Prioritize emerging **European models and datasets** for upcoming benchmarks.

- **Public Release & Collaboration:**
  - Publish results on **MINERVA github (**https://github.com/minerva4ai-eu/minerva-benchmarks/**)**.
  - Open-source selected tools and configs for community reproducibility.
  - Prepare materials for **T3.4 Best Practices documentation.**

Success Stories

Alexandros Paliouras

Success Story: SME Support

Initial objectives

- Mitigate random node failures when running pre-training in large scale (dozens of nodes)
- Debug and optimize code for large scale training of SLMs using NVIDIA Nemo and MegatronLM

Actions:

- Node health checks, failure tolerance of <5% (typical practice)
- Introduced NCCL specific setup for internode communication speedup & stability
- Fixed code bugs causing failures when scaling
- Conducted code profiling and scaling experiments

Results:

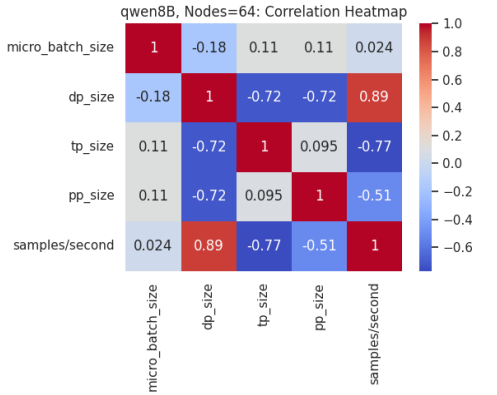
- Report and proposal of optimal parameters per model for MN5

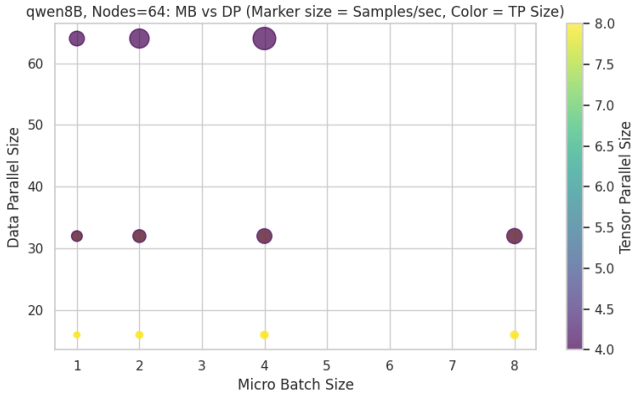
Success Story: SME Support

Our Feedback

```
def trainer_config():
    strategy = strategy_config()

    trainer = run.Config(nl.Trainer,

                         accelerator="gpu",
                         accumulate_grad_batches=1,
                         callbacks=callbacks_config(),
                         num_nodes=64,
                         devices=4,
                         limit_test_batches=30,
                         limit_val_batches=30,
                         log_every_n_steps=10,
                         max_epochs=1,
                         max_steps=1500,
                         plugins=bf16_mixed(),
                         strategy=strategy,
                         use_distributed_sampler=False,
                         val_check_interval=200,

                                 )
    return trainer
```

```
def trainer_config():
    strategy = strategy_config()
    plugins = [bf16_mixed()]
    trainer = run.Config(
        nl.Trainer,
        accelerator="gpu",
        accumulate_grad_batches=1,
        callbacks=callbacks_config(),
        num_nodes=int(os.environ["NUM_NODES"]),
        devices="auto",
        #limit_test_batches=30,
        limit_val_batches=LIMIT_VAL_BATCHES,
        log_every_n_steps=10,
        max_epochs=1,
        max_steps=MAX_STEPS,
        plugins=plugins,
        strategy=strategy,
        use_distributed_sampler=True,
        val_check_interval=VAL_CHECK_INTERVAL,

    )
    return trainer
```





Their Feedback

"Key Things We Learned
From your support, we now clearly understand:

- Why maximizing DP and minimizing PP/TP is usually the best strategy
- That our workload is compute-bound (micro-batch size barely affects throughput)
- How to properly integrate SLURM settings such as ntasks-per-node=4 and devices="auto"
- The importance of the batch size formula:
  global_batch_size = DP × micro_batch_size × accumulate_grad_batches"

✨ "...Overall, your guidance made a huge difference in our training stability, performance, and our understanding of how to get the most out of the system..."

✨ "...Once again, thank you for the fantastic support. The clarity of your explanations, the depth of your analysis, and how directly applicable everything was — all exceeded our expectations. We're looking forward to collaborating more as we scale up..."

Time to wrap up!

Claudio Arlandini

How to Get Support from MINERVA

- For support:
  - minerva4ai.eu/support/#support-form
  - ***support@minerva4ai.eu***

- A unified support team will bind expertise and resource overview across all the partners of the project which will act as common main entry for all the users.

- A multi-level, helpdesk support service will have a single point of entry for users, the AISC portal, backed up by a ticketing tool enabling user requests to be processed and the life of a ticket to be tracked through to resolution, while retaining its history.

- The service portal will be integrated with the ticketing system.

Final Discussion

🎤 **Let us know your point of view!**

- Have you got any question?
- Have you got any thought?

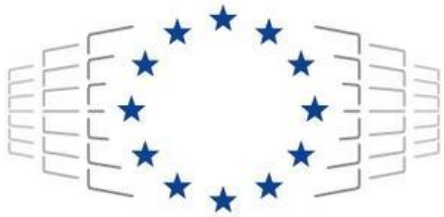
LinkedIn

Survey

Reach out to us @ *info@minerva4ai.eu*

Thank you

Co-funded by the European Union



EuroHPC Joint Undertaking

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101182737. The JU receives support from the Digital Europe Programme.

