

Project Application - [REDACTED]

The Project

Project details

Project title: Predictive Workload and Cooling Optimisation for Data Centres

Project summary (abstract):

GreenOpt AI, an SME based in Vienna, Austria, proposes to develop a forecasting system to improve energy efficiency in data centres. The system will predict cooling demand and workload distribution up to 30 minutes ahead, using telemetry data such as CPU/GPU utilisation, rack inlet/outlet temperatures, energy consumption, and weather conditions. By leveraging state-of-the-art architectures such as Temporal Fusion Transformers and PatchTST, alongside classical VARS, we will benchmark and optimise multivariate time-series forecasting approaches for scalable, distributed training on EuroHPC infrastructure. The outcome will be a production-ready forecasting model that reduces cooling over-provisioning, saving up to 15% in energy costs and lowering CO₂ emissions.

Keywords:

AI, time series forecasting, data centres, energy efficiency, workload prediction, cooling optimisation, PyTorch, DeepSpeed

Instructions: Not provided

Proposal for civilian purposes: true

Is any part of the project confidential?: No

Artificial Intelligence (AI) technology #1

AI set of technologies selection: Machine Learning

Share (%): 100

Application Domain #1

Application domain title: PE7 Systems and Communication Engineering

Application Domain share (%): 50

Application Domain #2

Application domain title: PE7 Systems and Communication Engineering

Application Domain share (%): 50

Submission details

Project duration: 3 months

Project Lead and Organisation information

Instructions: Not provided

Project Lead-Personal information

Gender: Female

Title: Dr.

First (given) name: Joanna

Last (family) name: Doe

Initials: JD

Date of birth: 03-11-1989 (CET)

E-mail address: joanna.doe@greenopt.ai

Secondary e-mail address: info@greenopt.ai

Nationality: Austrian

Phone Number: +43|012345678

Job title: CTO

Employment contract valid for more than 3 months after end allocation: true

Website: www.greenopt.ai

Organization details

Instructions: Not provided

Organization name: GreenOpt AI GmbH

Organization type: Startup

Company VAT number: ATU12345678

Organization with research activity: Yes

Organization head office is located in Europe: Yes

Organization department: R&D

Organization group: Not provided

Organization address: Reiverside Esplanade 1

Organization postal code: 1234

Organization city: Vienna

Organization country: Austria

Team Members Information

Does the project have Team Members: No

AI Factory Selection

AI Factory Selection:

Leonardo BOOSTER

Code(s) used: wandb, Optuna, DeepSpeed, Hugging Face Accelerate, Pytorch, Python

Requested amount of resources (GPU hours): 5000

Maximum number of GPUs: 8

Total storage required (GB): 2000

Total amount of data to transfer to/from (GB): 500

Code Details and Feasibility

Code details #1

Name and version of the code:

Python 3.10, PyTorch 2.3, PyTorch Lightning 2.2, Hugging Face Transformers 4.42

Webpage and other references:

www.huggingface.co

Description of the code:

Not provided

Scalability and performance

Describe the scalability and performance of the application:

The application is built for distributed training of multivariate time series models. Using PyTorch's Distributed Data Parallel and DeepSpeed ZeRO optimisation, it achieves near-linear scaling up to 2 A40 GPUs.

Optimization of the work proposed

Explain how the optimization work proposed will contribute to future large scale applications:

Optimisation focuses on profiling bottlenecks in I/O, communication overhead, and memory footprint. Results will guide best practices for future large-scale forecasting tasks in energy, climate, and industrial operations. Lessons learned will help scale applications to thousands of GPU-hours while maintaining efficiency and reproducibility.

Ethics Self-Assessment

Instructions: Not provided

Please specify how does your project ensure ethical principles and addresses potential societal impacts associated with the development and deployment of AI technologies:

The project complies with the EU's six ethical principles for trustworthy AI: respect for autonomy, prevention of harm, fairness, explicability, robustness, and accountability. The focus on sustainable infrastructure ensures a positive societal impact, lowering energy use and CO₂ emissions. No personal data is processed, and the AI system is used only for technical optimisation tasks.

Respect for Human Agency

Please describe how your system ensures that end-users have the ability to control vital decisions about their own lives:

Final decisions about workload and cooling remain with data centre operators. The AI provides forecasts and recommendations, but operators retain full control. Human-in-the-loop governance ensures vital decisions always rest with people.

Privacy & Data Governance

Does your proposal involve handling of personal data?: No

Please describe how data is collected and processed from the aspect of lawfulness, fairness and transparency:

None. Only technical telemetry data is processed.

What measures (such as anonymization, pseudonymisation, encryption, and aggregation) you took to safeguard the rights of data subjects?:

None. Only technical telemetry data is processed.

Please describe the measures you employ to prevent data breaches and leakages:

None. Only technical telemetry data is processed.

Fairness

How do you ensure avoiding algorithmic bias, in input data, modelling and algorithm design?:

Not applicable.

Individual, and Social and Environmental Well-being

If relevant, describe how the AI system is mindful of all stakeholders and the environment:

The system directly reduces CO emissions by up to 15% in data centre cooling, contributing to EU climate goals. Stakeholders benefit through lower operational costs and sustainable operations.

Transparency

Are the end-users aware that they are interacting with an AI system?: Yes

Please describe how the participants and/or end-users will be informed about interacting with an AI system, and about its purpose, capabilities, limitations, benefits and risks:

End-users are fully aware they are interacting with an AI-assisted system. Documentation and interfaces communicate the purpose, benefits, limitations, and risks clearly.

Accountability

Please describe how your system ensures that potential ethically and socially undesirable effects will be detected, stopped, and prevented from reoccurring:

System outputs are logged and monitored. Unexpected or undesirable effects (e.g. anomalous cooling recommendations) are flagged automatically and reviewed by human operators. Corrective actions prevent recurrence.

Data Consent

Eligibility confirmation

Instructions: Not provided

I confirm that the Project Lead and all Team Members are affiliated with organizations listed on the Horizon 2020 countries list.: true

I confirm that the submitted information is correct and complete and comply for the entire duration of the action with the eligibility criteria set out in the call conditions.: true

divider: Not provided

Data consent

Instructions: Not provided

space: Not provided

In case the proposal is awarded, EuroHPC JU would like to publish the Project Lead and Team Members' names and organizations. This may involve sharing this information on our website, social media channels, or in other promotional materials related to the project. Please provide your consent below.: I consent

In order to submit the proposal, you must accept the terms and conditions stated in the Access Policy, hence confirming that you have read and understood the call procedures. The documentation can be found at : I accept the terms and conditions stated in the Access Policy