Accelerating Generative AI with PyTorch

Introduction to the HPC System

Meet the Team

Constantine Dovrolis

Christodoulos Stylianou

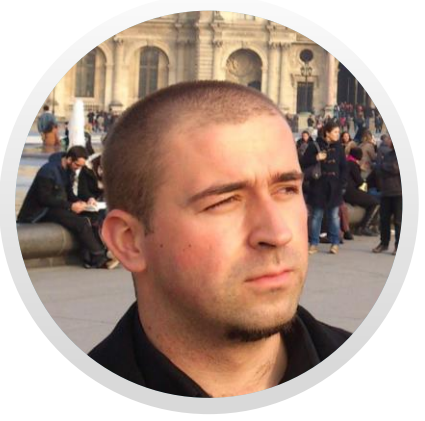
Robert Jan Schlimbach

Sophia Honisch

Ivan Gentile

Charalambos Chrysostomou

Boris Velichkov



Miriam Koch

Presenters

- Boris Velichkov
  - Lead Presenter (Introduction and Setup) and Instructor (Practical Session)
  - Researcher for the EuroCC2 Project at Sofia University
  - boris.velichkov@fmi.uni-sofia.bg
- Robert Jan
  - Lead Presenter (PyTorch Intro and Profiling) and Instructor (Practical Session)
  - Technical Advisor ML/HPC at SURF BV
  - robert-jan.schlimbach@surf.nl

Introduction to the HPC System

- Introduction
  - GWDG HPC System
  - Setting Up HPC Environment
- PyTorch and Profiling
  - Learn how to train machine learning models using PyTorch and optimize their performance with profiling on the provided HPC system
- Practical Session
  - Gain hands-on experience with PyTorch and profiling on the HPC system as you work through guided exercises with expert support

The GWDG at a glance

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

The computer center and IT competence center for the Georg-August-University Göttingen and the Max Planck Society.

Central areas of responsibility

- Modern and secure IT infrastructure
- IT support for excellent research
- In-house research for innovative IT services

Various supra-regional tasks:

- National High Performance Computing Center
- National HPC Center of the DLR
- AI service center for sensitive and critical infrastructures
- Data center in four NFDI consortia
- Host for DARIAH-EU, German National Library, GFBio, NUM CODEX, MWS, WirLernenOnline, etc.
- Cloud operator, including Academic Cloud for universities in Lower Saxony

KISSKI: AI Service Center for Sensitive and Critical Infrastructures

Research into AI methods and their provision in a highly available AI service center for critical and sensitive infrastructures. Focus on socially highly relevant fields of medicine and energy

Service offer:

- Infrastructure
  - Hardware
  - Software
  - Models & data
- Consulting
  - Initial consultation
  - Further consultation
- Development
- Training
  - Training courses designed specifically for KISSKI
  - External courses offered by the KISSKI consortium partners

KISSKI
KI-Servicezentrum für sensible und kritische Infrastrukturen

NHR-NORD@GÖTTINGEN

- As a high-performance computing center, the GWDG, together with the University of Göttingen, is now one of nine members of the National High Performance Computing Network - NHR. It currently operates the NHR systems Emmy and Grete.

- Competencies
  - Life sciences
  - Earth system sciences
  - Fluid mechanics
  - AI and big data
  - Digital humanities

- Access to the NHR systems
  - The allocation of computing time for small (test) projects is informal
  - A short application is required for extensive projects https://docs.hpc.gwdg.de/application_process/index.html
  - Approved projects can be managed in our convenient portal https://hpcproject.gwdg.de

Become customers at NHR-NORD@GÖTTINGEN & KISSKI

NHR-NORD@GÖTTINGEN (https://gwdg.de/en/community-pages/nhr-intro/)

To be able to use the computing systems, you must join an existing project or apply for a project. The following options are available:

- Join an existing project
- Apply for test access with limited computing time (requirement: member of a German university)
- Apply for a full project (requirement: member of a German university)

KISSKI (https://kisski.gwdg.en/)

To book services of KISSKI an AcademicID is required. The following options are available:

- Federated login: log in with your own institution's access data
- New registration: create an AcademicID during the booking process

Acknowledgments

Acknowledgment

- Funded by the European Union, as part of the EuroCC2 Project, this work has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101101903. The JU receives support from the Digital Europe Programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia.
- We acknowledge the support of the Academic Cloud and the Georg-August-University Göttingen (GWDG) for providing the HPC System.

Disclaimer

- Funded by the European Union, as part of the EuroCC2 Project, the views and opinions expressed in this work are solely those of the author(s) and do not necessarily reflect those of the European Union or the European High-Performance Computing Joint Undertaking (JU) and participating countries in the project. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by the European Union

EuroHPC Joint Undertaking

EURO²

ACADEMIC CLOUD

GWDG Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Setting Up Your HPC Environment

- Log in to provided accounts
- Access the shared folder and SSH key
  - https://owncloud.gwdg.de/index.php/s/2Ewlpc26GZujGC0
- Access the GitHub repository
  - https://github.com/CaSToRC-CyI/isc2024-tutorial
- Go to the introduction README.md file
  - **The GitHub repository → intro → README.md** (https://github.com/CaSToRC-CyI/isc2024-tutorial/blob/main/intro/README.md)
- Follow the steps in the README.md file and set up your environment
  - Install SSH clients (if needed)
  - Configure SSH
  - Log in to the cluster
  - Run the introduction Jupyter Notebook
  - Set up SSH Tunnel
  - Visualize the notebook in your local browser