Luigi Pagani

Portfolio Mobile: +39 3404896210 Email: luigi2.pagani@mail.polimi.it GitHub: Luigi Pagani LinkedIn: Luigi Pagani

Summary

Machine Learning Engineer with expertise in High-Performance Computing (HPC) and deep learning frameworks. Experienced in building end-to-end ML pipelines and optimizing training and inference through distributed and multi-GPU systems. Skilled in Large Language Model (LLM) training, deployment, and MLOps best practices. Background includes neural network research and applications in computational engineering and digital twins obtained through graduate projects, master thesis and internship.

Professional Experience

Machine Learning Engineer Apr 2025 – Present
Nebul, Leiden, Netherlands
- Focused on LLM inference deployment on GPU-backed **Kubernetes** clusters using **vLLM**.
- Implemented LLMOps and GitOps workflows with **Helm** templating for one-click deployments.
- Developed one-click benchmarking of dozens of vLLM model configurations and GPU flavors.
- Reduced deployment time by **80**%.

ML Research Engineer Intern Oct 2024 – Mar 2025
Siemens Digital Industries Software, Leuven, Belgium
- Developed transformer-based neural networks for fast numerical PDE solvers, addressing unstructured meshes and time-dependent boundary value problems.
- Implemented **PyTorch DDP** (Distributed Data Parallel) for multi-GPU training, enabling large-scale experimentation.

Individual Contributor Project Numina Aug 2024 – Jan 2025
Remote
- Work contributed directly to the paper highlighted in the following section.
- Built an LLM pipeline for automatic evaluation, achieving a **300% speedup** compared to previous solutions.
- Created a synthetic data generation system for high school math problems using open-source LLMs, improving dataset scalability.
- Developed a bootstrapping pipeline to auto-formalize natural language into **Lean 4** statements, leveraging **vLLM** & **LLaMA-Factory**.

Papers

Kimina-Prover Preview: Towards Large Formal Reasoning Models with Reinforcement Learning Apr 2025
Related to Project Numina
ArXiv: 2504.11354

Education

MSc in High Performance Computing Engineering Mar 2023 – Mar 2025
Politecnico di Milano, Italy
Grade: 110/110, *cum laude*
Recipient of merit-based scholarship for outstanding academic performance

BSc in Mathematical Engineering Sep 2019 – Sep 2022
Politecnico di Milano, Italy
Final Grade: 103/110

Technical Skills

Programming: Python, C/C++, CUDA

Infrastructure: Docker, Kubernetes, Helm templating, GitOps

ML Libraries: PyTorch, vLLM, LiteLLM, TensorFlow

Certifications: AWS SAA-C03, Azure AZ-104, CKA