Luigi Pagani

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

Summary

ML Engineer specialized in LLMs and deep learning. Experienced building end-to-end ML pipelines, optimizing training/inference on distributed, multi-GPU systems. Strong in LLM training, evaluation and deployment, RL environments development and MLOps best practices.

Professional Experience

Machine Learning Engineer Apr 2025 – Present
Nebul — Leiden, Netherlands

- Built **LLMOps/GitOps** workflows with **Helm** for one-click, reproducible deployments and benchmarking on Kubernetes.
- Co-developed a production **LLM inference API** on a 120-GPU **Kubernetes** cluster using **SGLang** and **vLLM**, managed via Open Model Engine CRs
- Operationalized NVIDIA's cloud-native stack: **GPU Operator** (drivers, device plugin, Container Toolkit, DCGM) and **Network Operator** (RDMA/GPUDirect, CNI) for high-throughput, low-latency serving.

ML Research Engineer Intern Oct 2024 – Mar 2025
Siemens Digital Industries Software, Leuven, Belgium

- Developed transformer-based neural networks for fast numerical PDE solvers on unstructured meshes and time-dependent simulations.
- Implemented **PyTorch DDP** (Distributed Data Parallel) for multi-GPU training.

Individual Contributor — Project Numina Aug 2024 – Jan 2025
Remote

- Built an automated LLM evaluation pipeline for high-school math problems using the **OpenAI Batch API** for verification and **vLLM** for rollouts generation.
- Designed a synthetic data generation pipeline for math problem dataset creation with open-source LLMs.
- Developed a bootstrapping pipeline to auto-formalize natural language into **Lean 4** statements, and fine-tuning on them with **LLaMA-Factory**.

OSS Contributions

Prime Intellect — Environments Hub Sep 2025

- Ported AidanBench, a creativity & long-context RL/evaluation benchmark, to the verified Prime Intellect Environments Hub. Link

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, Go & C/C++ (familiarity)

Infrastructure: Docker, Kubernetes, Helm, GitOps, Argo CD

ML Libraries: PyTorch, vLLM, SGLang

LLM Stack: LangChain, Langfuse, LiteLLM, Verifiers