

Cristian C. Spagnuolo

NLP ENGINEER

☎ (+39) 3395869455 | ✉ cristian.c.spagnuolo@gmail.com | 🏠 cris96spa-latent-space.com | 📱 cris96spa | 🌐 cristian-c-spagnuolo

Summary

NLP engineer on production LLMs: quantization and serving-runtime optimization, fine-tuning (SFT, DPO, GRPO), and evaluation pipelines with non-regression tests. Playing 12 to 14 hours a day, and I have not yet found the part that feels like work.

Experience

Artificially SA

Lugano, Switzerland

NLP ENGINEER

05/2025 - Present

- Optimized on-premise LLM inference with quantization  and **TensorRT/vLLM** runtimes: **2x less GPU memory, 1.8x faster**.
- Drove the internal **fine-tuning framework (SFT, DPO, GRPO)** as main contributor: multi-GPU training and **MLflow** tracking.
- Built an **LLM evaluation pipeline** (retrieval, metrics, artifacts, non-regression tests), cutting **eval effort by 66%**.
- Engineered a **tabular-data Q&A agent** with secure code generation, visualization, and forecasting; **>90% accuracy**.
- Authored an **automatic API-doc generator** for the internal Python template, generating **MkDocs** pages from source on every build.

ML cube

Milan, Italy

AI/ML ENGINEER

09/2024 - 05/2025

- Trained a **pricing agent** with PPO for hotel booking, increasing simulated revenue by **16%** over baseline strategies.
- Created an **LLM-powered SQL agent** for financial users, enabling **natural language queries** with **91% accuracy**.
- Designed an **AI regulatory assistant** on knowledge graphs, improving **analyst decision accuracy by 38%**.
- Extended the “ML cube Platform” with segmented drift-detection monitoring, raising **detection coverage by 35%**.
- Automated **document digitization** through a pipeline for structured data extraction, achieving **87% OCR accuracy**.

Education

Politecnico di Milano

Milan, Italy

MASTER'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING - **110/110 CUM LAUDE**

02/2020 - 07/2024

Università degli Studi del Sannio

Benevento, Italy

BACHELOR'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING - **110/110 CUM LAUDE**

09/2015 - 10/2019

Projects

latent-space

PROJECT

07/2026

- Built a personal site that runs a real **GPT-2 124M** forward pass - tokenize, prefill, per-token decode - streaming the bio token by token.
- Designed a **FastAPI** content backend with validated, sanitized Markdown rendering and a **React/TypeScript** frontend.
- Technologies:** Python, FastAPI, Pydantic, React, TypeScript, Vite, Tailwind, Docker, Vercel

fyt - From YAML Training

PROJECT

10/2025

- Authored a **configuration-driven ML library** for tabular tasks: define, train, tune, and reuse full pipelines from one YAML file or from Python.
- Implemented **leakage-safe** tuning, re-fitting preprocessing and feature selection inside every **Optuna** cross-validation fold.
- Technologies:** Python, Pydantic, Polars, scikit-learn, XGBoost, Optuna, MLflow, uv

Teleport MDPs: Curriculum Learning for RL

Advisor: Prof. Marcello Restelli

MASTER'S THESIS

02/2023 - 07/2024

- Formalized **Curriculum Learning** for **Reinforcement Learning** via the **Teleport MDP**: a teleport rate truncates the planning horizon, making a curriculum an annealing schedule.
- Derived theoretical bounds and an exact algorithm, plus two practical schedules that beat vanilla baselines on sparse-reward benchmarks.
- Technologies:** Python, PyTorch, NumPy, Gymnasium, PPO

Skills

Programming

Python, Java, C, C++, Assembly, SQL

AI/ML

vLLM, TensorRT, PyTorch, Hugging Face, MLflow, TensorFlow, LangChain, LangGraph, Ollama, OpenAI, Mistral, Qwen, DeepSeek, Gemini, Vertex AI, Pandas, Polars, NumPy, scikit-learn, Gymnasium, Matplotlib, Plotly

Technologies

GCP, AWS, Azure, Git, GitHub, Docker, Docker Compose, Traefik, MongoDB, FastAPI, Firebase, Django, Node-RED, Make, uv, Neo4j

Languages

English, Italian

Achievements

COMPETITIONS

2025 **2nd Place**, Hackapizza 

IBM Studios, Milan

OPEN SOURCE CONTRIBUTIONS

2026 **huggingface/trl**, Packing and padding-free SFT for VLMs on text-only data 

2026 **swiss-ai/vllm**, Apertus 1.5 loads standard and quantized checkpoints 