Davide Modolo

AI ENGINEER | MSC ARTIFICIAL INTELLIGENCE

davide.modolo98@gmail.com
linkedin.com/in/davide-modolo
github.com/davidemodolo

Education

University of Trento

Trento, Italy

Master's Degree in Artificial Intelligence Systems

September 2021 - March 2025

- **Thesis**: Exploring the Use of LLMs for Agent Planning: Strengths and Weaknesses. *[report | code]*
- **Relevant Courses**: Machine Learning, Deep Learning, Natural Language Understanding, Automated Planning, Law & Ethics in AI, Fundamentals of AI (Reasoning and Planning), Signal Processing, AI for Finance, HPC.

University of Trento

Trento, Italy

Bachelor's Degree in Computer Science

September 2017 - June 2021

- **Thesis**: Healthy Plus - Redesign and evolution of an Android application for monitoring healthy lifestyles.
- **Core Topics**: Algorithms & Data Structures, Software Engineering, Database Systems, Operating Systems, Computer Architectures, Distributed Networks.

Work Experience

OpenCity Labs - *Startup*

Remote

AI ENGINEER - R&D

August 2025 - Current

- Architected a multi-tenant conversational AI system using the open-source Cheshire Cat AI framework, developing and releasing multiple open-source plugins; managed Gemini API integrations while benchmarking Docker/Grafana infrastructure to self-host embedding models.
- Started the adoption of the Model Context Protocol (MCP) to standardize tool execution; engineered a custom Client-Server POC for autonomous appointment booking, establishing the architectural pattern to expose future company microservices as modular AI endpoints.
- Engineered a core NLP pipeline, fine-tuning multilingual Transformer-based NER models for PII anonymization; built a spaCy-based sentiment analysis module optimized for fast CPU inference to monitor user messages, and implemented an output validation layer for automated translation.

`Python, NLP (spaCy/Transformers), MCP, Gemini API, Open-Source Embeddings, Docker, GitHub, Grafana, Prometheus`

Eurecat Technology Center

Barcelona, Spain

INTERN / AI ENGINEER

April 2024 - June 2024

- Implemented a token-level uncertainty estimation framework for planning tasks by analyzing log-probabilities, benchmarking performance across proprietary (GPT-4o) and locally quantized models (Llama 3.1 via llama.cpp).
- Developed a RAG system utilizing *gte-large* embeddings to retrieve context-aware user preferences, significantly reducing model hallucination in recurrent decision-making scenarios.
- Prototyped a multimodal agent integrating LLaVA for visual reasoning tasks and validated the uncertainty reconstruction approach across multiple simulated environments.

`Python, PyTorch, Hugging Face, OpenAI API, llama.cpp, Ollama`

Projects

COVID-19 Lung Ultrasound Images classification - Medical Imaging Diagnostic

[slides | report | code]

- Designed a Multi-stage deep learning model to classify Lung Ultrasound images based on a multiclass severity scoring (0-3).
- Developed the 3 components of the model: a multi-class frame classifier, an uncertainty detection model and a similarity module.
- Dataset composed of anonymized images from 14 real patients for a total of 47k frames used to train and fine-tune various Computer Vision models.

`Python, PyTorch, scikit-learn, Pandas, NumPy, Azure, Computer Vision`

Joint Intent Detection and Slot Filling - Natural Language Understanding

[report | code]

- Implemented 4 Deep Learning models to simultaneously learn Intent Detection and Slot Filling tasks.
- Reached the goal to improve consistently by at least 2% over the base results on the ATIS dataset in both tasks, improved the baseline by 13% on the SNIPS dataset in the slot filling task.
- Fine-tuned BERT and ERNIE; built from scratch a model based on Bidirectional LSTM and an Encoder-Decoder architecture.

`Python, PyTorch, NLP (BERT/ERNIE/LSTM)`

Domain Adaptation / Transfer Learning - Deep Learning - *Team Project*

[code]

- Built, trained and evaluated a deep learning model on a standard setting of Unsupervised Domain Adaptation.
- Solution based on ResNet34 plus a custom adaptation layer with loss based on direct 3rd-order and grouped 4th-order statistics.
- Approach tested on the Adaptiope dataset, reaching a +11.53% improvement with respect to the baseline non-adapted model.

`Python, PyTorch, Google Colab`

Skills

Languages	`Python, JavaScript, C/C++, SQL, Bash, LaTeX`
ML & Data	`PyTorch, Hugging Face, TensorFlow, Transformers, spaCy, scikit-learn, Pandas, NumPy, OpenCV`
DevOps & Tools	`Git, Docker, Azure, Grafana, Prometheus, MPI (High Performance Computing)`
Spoken Languages	**English** - Professional, **Italian** - Native