Alberto **Formaggio**

DATA SCIENTIST

📱 (+39) 3457045718 | ✉ alb.formaggio@gmail.com | 🏠 albertoformaggio1.github.io | 💻 Albertoformaggio1 | 💼 alberto-formaggio

Education

Artificial Intelligence and Robotics - Master's Degree *Padua, Italy*

University of Padova *September 2022 - December 2024*

- Grade Point Weighted Average: 30/30 (**4.0 GPA**). Top 1% of the study program.
- **Main courses:** Computer Vision, Operations Research, Learning from Graphs, Machine Learning.

Computer Science and Artificial Intelligence - Exchange Semester *Prague, Czech Republic*

Charles University *February 2023 - July 2023*

- **Main courses:** Deep Learning, Natural Language Processing, Speech Recognition, Data Mining, Multiagent Systems.

Computer Engineering - Bachelor's Degree *Padua, Italy*

University of Padova *September 2019 - July 2022*

- Final Grade: **110/110 summa cum laude**. Top 1% of the study program (4.0 GPA).
- **Main courses:** Android Development, Software Engineering, Database Design, Linear Optimization.

Work Experience

NLP Master Thesis *Mannheim, Germany*

ABB Research Center *May 2024 - October 2024*

- Converted 4000 unstructured technical reports into a **Neo4J knowledge graph** by leveraging pre-trained **Large Language Models**.
- Designed and developed interactive dashboards to engage with the generated Knowledge Graphs, serving technicians across 20 countries.
- Leveraged Knowledge: Applied Pandas, Langchain, Ollama, and Azure OpenAI for data manipulation and extraction with LLMs; programmed in Python; and employed Neo4J and Cypher for storing and querying the knowledge graph.

Computer Vision Engineer Intern *Mannheim, Germany*

ABB Research Center *March 2024 - April 2024*

- Created a fine-tuning pipeline for an industry-specific **object detection** task, optimizing performance for high-resolution images.
- Established a strong baseline with validation mAP of 0.8 by using deep learning models like RetinaNet.
- Leveraged Knowledge: Employed PyTorch for data preprocessing and built a modular pipeline with PyTorch and PyTorch Lightning, integrating tracking capabilities with MLflow.

Projects

An Empirical Study on Ensemble of Segmentation Approaches *Paper published in Signals*

MATLAB for Deep Learning *June 2022*

- Accomplished an 89.3% Dice score in **image segmentation** by using diverse ensembles, surpassing previous methods.
- Contributed to the paper by enhancing object border detection accuracy by 1% (from 90.7% to 91.7%) through the implementation of new loss functions [**Code**, **Paper**].

Better Quality Embeddings for Node Classification Combining Classification with Link Prediction in Graph Neural Networks

PyTorch, Pytorch Geometric, Python

- Implemented link prediction as a self-supervised pre-training technique in **Graph Neural Networks** to enhance node classification performance.
- Reached 0.8 accuracy on the Cora dataset by evaluating GCN, GAT, and GraphSAGE with MLP augmentation [**Code**, **Report**].

Twitter Sentiment Analysis: Public Perception of Climate Change

Tensorflow, Pandas, Scikit-Learn, Numpy, Python

- Achieved an 86% F1-Score in climate change **sentiment analysis** on Twitter using preprocessing, word embeddings, class balancing, and machine learning techniques (e.g. SVM, BERT) [**Code**, **Blog Article**].

End-to-End Retrieval Augmented Generation (RAG) pipeline

Langchain, Ollama, Streamlit, Python

- Extracted datasets through web scraping using BeautifulSoup and created a **Retrieval-Augmented Generation (RAG)** chain with Ollama.
- Implemented a web interface for querying the LLM and retrieving answers from 224 documents by using Streamlit [**Code**, **Blog Article**].

Honors & Awards

4 x Merit Scholarship Winner, University of Padova *Padua, Italy*

Granted four €1,000 merit-based scholarships, awarded to top students at the University of Padova.