MATTIA CURRI

Master's Degree Computer Science Student - Software Developer - AI Engineer

mattia2002@gmail.com | mattiacurri | mattiacurri | mattiacurri.github.io

ABOUT ME

Ciao! I'm Mattia, an Italian Computer Science bachelor graduate. I'm interested in expanding my cultural boundaries, improving my technical and professional knowledge I am willing to work in the software development field, although I have little experience, I have a lot of interest and willingness to learn new things.

EDUCATION

Master's Degree in Computer Science - Artificial Intelligence Track — **University of Bari**

Current GPA: 4.0/4.0 — Sep 2024 - current

- Course entirely taught in English
- Relevant courses taken: Numerical Methods, Database II, Formal Methods, Fundamentals of Artificial Intelligence, Natural Language Processing, Software Engineering for AI-Enabled Systems

Bachelor's Degree in Informatica (Computer Science) — **University of Bari**

110 with honours/110 (Final GPA: 3.8/4.0) — Sep 2021 - Jul 2024

- Thesis in Computational Intelligence: *Supervised Learning Techniques for Semantic Segmentation of Aerial Images of Agricultural Fields*
  - 3 Months research internship at the Cilab Lab of the University of Bari under the supervision of Prof. Gennaro Vessio.
  - Comparison of Convolutional and Transformer architectures to tackle the Supervised Task of the Agriculture-Vision Challenge at CVPR 2024, training models on ~ 50k images.
- Relevant courses taken: Computational Intelligence, Algorithms and Data Structures, Databases, Software Engineering, Knowledge Engineering, Information Retrieval, Human-Computer Interaction, Computer Networks.

PROJECTS

Code Comment Classification System - Team Project — **Software Engineering for AI Exam Project**

`Python` CatBoost `Grafana` `FastAPI` `Docker` `MLflow` `Prometheus` `Github Actions` — CodeCommentClassification

- Architected a production-grade multi-label classification system categorizing code comments across Java, Python, and Pharo
- Engineered end-to-end ML pipeline: data versioning (DVC) → automated preprocessing → dual-model training (CatBoost + SetFit) → experiment tracking (MLflow) → containerized deployment
- Implemented comprehensive data quality framework improving dataset integrity by 12%
- Deployed full-stack solution: **FastAPI** backend with async processing → **Gradio** frontend → **Prometheus**/**Grafana** monitoring, with sub-2s inference latency
- Achieved 92.7% test coverage across unit, integration, and behavioral test suites, ensuring production reliability
- Published live demo on **Hugging Face Spaces** with automated CI/CD, demonstrating MLOps best practices for academic coursework exceeding industry standards

EmPULIA Knowledge Graph-Based RAG Approach for Question Answering - Solo Project — **AI/NLP Project**

`Python` `Knowledge Graphs` `RAG` `Neo4j` `FAISS` `Transformers` Ollama — EmPULIA System

- Developed a comprehensive **Retrieval-Augmented Generation** system analyzing $\sim$ 20 legal documents (up to 200 pages each) through an end-to-end pipeline
- Built an end-to-end pipeline: web scraping → PDF processing → knowledge graph extraction with **Gemini** → semantic indexing → RAG inference
- Implemented **knowledge graph normalization** using DBSCAN clustering and semantic embeddings to reduce entity redundancy
- Created custom evaluation framework with **Context Faithfulness**, **Context Precision/Recall**, and **Answer Accuracy** metrics
- Integrated multiple retrieval strategies: standard, multi-query, entity extraction, and Neo4j random walks

Temporal Link Prediction on Social Networks - Solo Project — **BigData Exam Project**

`Python` `PyTorch` NetworkX Weights & Biases — Temporal Link Prediction

- Developed a **temporal link prediction** system using **Evolving Graph Convolutional Networks (E-GCN)** to forecast future social connections in the **Gab** network
- Integrated **BERT embeddings** (768-dimensional) to capture semantic content from user posts and enhance prediction accuracy
- Engineered a multi-snapshot temporal pipeline handling **6 historical time steps** with incremental training and fine-tuning capabilities
- Achieved **MAP=0.87**, **Macro F1=0.86**, **AUC-ROC=0.80** on temporal link prediction tasks
- Conducted experiments on synthetic data generation and evaluated model generalization across different network structures

Corporate Credit Rating Prediction - Solo Project **AI Project**

</> **Python** **Scikit-learn** **Pandas** **Pgmpy** Imblearn Matplotlib Numpy Rating Prediction

- The goal of the project is to successfully predict the rating assigned by rating agencies to a company
- Dealed with an unbalanced dataset using **Class Weights**, **SMOTE**, **ADASYN**, **SMOTETomek**, **SMOTEENN**
- Implemented a **Bayesian Network** to do probabilistic queries, with a correlation score based on *Balanced Accuracy* of **0.59**
- Achieved up to **86%** of **Balanced Accuracy** using XGBoost and Random Forest

PUBLICATIONS

A. Porcelli, F. Di Gravina, E. Fontana, **M. Curri**, F. D. Di Gregorio. *"FFT-UniBa at Cruciverb-IT: Special Length Tokens and CSP for Italian Crossword Solving."* In **EVALITA 2026**, Pre-print: https://apa.dipsco.unitn.it/evalita2026/45.pdf

- Fine-tuning of Italian T5 with length-constrained generation and CSP-based grid solver; MRR 0.63, 34% full grid accuracy.

AWARDS, HACKATHONS & COMPETITIONS

First Italian University AI Competition

</> **AI** **LLM** Open Data February 2026

- **Achieved 2nd place** in the first stage of the First Italian University AI Competition organized by AI2B, out of 20 teams.
- The competition involved using open data to create a solution for the city of Bari, specifically focused on the problem of logistics and intelligent traffic management.

ITAData Hack 2025

</> **Python** **CatBoost** **Scikit-Learn** Imblearn Pandas Sept 2025

- Anomaly Detection task in Hadoop Distributed File System Logs.
- **Secured 2nd place in the ITAData Hack 2025 competition** out of 15 teams.

AI2B Hackathon Winner

</> **Cryptography** **Prompt Engineering** June 2025

- Winning team of the AI2B Hackathon organized by University of Bari and AI2B, focused on AI and cybersecurity.
- Obtained €500 winning first place out of 10 teams.

CyberChallenge 2024

</> Web Binary Exploitation Cryptography Network Mar 2024-July 2024

- Qualified to the **national stage** by ranking in the top 6 of the local venue, reaching 21st place out of 43 teams at the national competition.
- Qualified for the local stage by being in the top 20 of the selection.

ADDITIONAL

Languages: Italian (native), English (B2 Cambridge)