

Yahya Momtaz

📍 Italy

✉️ yahyamomtaz@gmail.com

🔗 github.com/yahyamomtaz

🔗 <https://yayamomt.tech>

🔗 linkedin.com/in/yahya-momtaz-601b34108

📄 PerMESSO DI SOGGIORNO: Lavoro subordinato (Italia)



Professional Experience

AI Researcher

05/2024 – present | Naples, Italy

University of Naples Federico II

- Currently developing an automated data management pipeline that extracts text, utilises LLMs for data cleaning and summarisation, and implements Retrieval-Augmented Generation (RAG) to provide context-aware updates to the project website.
- Applied deep learning and image processing techniques to restore historical manuscripts, reducing bleed-through and improving readability.
- Engineered an OCR pipeline for early printed books using Kraken and PyTorch; fine-tuned a ByT5 language model to perform post-OCR correction, to increase text extraction accuracy.
- Developed a full-stack web application for bleed-through removal—built the backend with Flask, the frontend with React, containerised it with Docker, and deployed it on Google Cloud Platform (GCP).
- **Demo:** <https://bleed-through-cleaner-app-469035949643.europe-west1.run.app/>
- Developed the MAGIC project website using React, implementing CI/CD automation pipelines with GitHub Actions for production releases. <https://www.magic.unina.it/>
- Authored research in J. Imaging and J. Electronics and presented technical findings at international conferences, including AIUCD and RAD 2025.

Data Scientist (internship)

01/2022 – 08/2022 | Naples, Italy

Spici srl

- Developed a data-driven web scraper to collect and structure data from LinkedIn, automating dataset generation for marketing analysis, ensuring proper data management and quality, and integrating with GCP data warehouse solutions to create efficient data pipelines.
- Implemented a Flask API for the Neo4j database, automating common tasks and reducing graph database implementation time by 40% for the team.

Network Engineer

06/2018 – 11/2020 | Tehran, Iran

Saba INV

- Maintained and monitored the company's internal network infrastructure, ensuring reliable connectivity and system uptime.
- Configured and managed servers, routers, and network devices to support day-to-day business operations.
- Installed and implemented hardware solutions for employees, improving workstation performance and network access.

Education

Master in Data Science

10/2021 – 03/2024 | Naples, Italy

University of Naples Federico II

Apple Developer Academy

09/2022 – 06/2023 | Naples, Italy

Bachelor in Software Engineering

09/2011 – 09/2016 | Karaj, Iran

Karaj Azad University

Certificates

<https://www.linkedin.com/in/yahya-momtaz-601b34108/details/certifications/>

Skills

Python

PyTorch, TensorFlow, Scikit-Learn, Hugging Face, NLP (LLMs, RAG, Text Mining), Model Deployment, Fine-Tuning

OpenCV, YOLOv8, Image Segmentation

Pandas, NumPy, Matplotlib, Plotly, Tableau

SQL, Neo4j, Data Pipelines, Apache Airflow

Docker, Git, GitHub Actions (CI/CD), Google Cloud Platform (GCP), AWS (Familiar)

React, JavaScript

Languages

English: C1 • **Italian:** B1

Projects

AI Image API

- Developed and deployed a real-time API for object detection and image enhancement using *YOLO and PyTorch, including API integrations for seamless data exchange.*
- Built the backend with *Flask*, containerized with *Docker*, and deployed on *GCP*.
- Created a responsive *JavaScript frontend* for live interaction.

Demo: <https://aiimagelive.web.app/>

Multi Agent TradeGraph

- Contributed to a multi-agent AI financial advisor using LangGraph, OpenAI GPT-4, and Firecrawl for real-time stock analysis and portfolio optimization.
- Resolved critical configuration and validation issues, enabling end-to-end analysis pipeline for multi-stock portfolio recommendations

Repository: <https://github.com/ApexAILabs/TradeGraph>

Retrieval Augmented Generation (RAG) Chatbot:

- Developed a *document-based QA chatbot* for Italian legal texts, combining *LLMs + IR techniques* for accurate, context-aware answers.
- Implemented backend in *Python*, web interface with *Streamlit*, and integrated text-mining algorithms.

Repository: <https://github.com/yahyamomtaz/ChatBot>

Prompt Analyzer

- Developed a *Streamlit app* for interactive dataset exploration with *LLM-powered insights*.
- Focused on industrialization of the Prompt Analyzer by creating CI/CD pipelines using GitHub Actions and *implementing MLOps best practices*.
- Integrated OpenAI API for prompt-driven analysis.
- AI Skills: Prompt Engineering, Context Engineering*

Repository: <https://pandasai-dataset-analysis.streamlit.app/>