



Yahya Momtaz

📍 Italy 🔗 <https://yayamomt.tech> ✉ yahyamomtaz@gmail.com

🌐 [linkedin.com/in/yahya-momtaz-601b34108](https://www.linkedin.com/in/yahya-momtaz-601b34108) 🐙 github.com/yahyamomtaz

Professional Experience

Researcher (Data Science & Machine Learning),

05/2024 – present | Naples, Italy

University of Naples Federico II

- Built an automated ETL pipeline using Docker and Apache Airflow to automate workflow, orchestrating parallel data extraction, database updates, and static file generation for a digital library platform.
- Applied model development techniques to fine-tune a post-OCR language model with PyTorch, using applied machine learning and generative AI to improve text extraction operation.
- Applied deep learning and image processing techniques using machine learning, Python, and TensorFlow to restore over 7,500 historical manuscripts, performing data analysis and feature engineering to reduce bleed-through and improve readability.
- Developed a full-stack web application for bleed-through removal, using Flask and React, containerized with Docker, and deployed on Google Cloud Platform (GCP) following agile software development methodologies.

Demo: <https://bleed-through-cleaner-app-469035949643.europe-west1.run.app/> 🔗

- Launched the project website (developed the frontend with React): <https://www.magic.unina.it/> 🔗

Data Scientist (internship), Spici srl

01/2022 – 08/2022 | Naples, Italy

- 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, Saba INV

06/2018 – 11/2020 | Tehran, Iran

- 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, University of Naples Federico II

10/2021 – 03/2024 | Naples, Italy

- Thesis: Enhancing object detection in low-light conditions using fine-tuned YOLOv8.
- Link to one of my presentations: <https://www.youtube.com/watch?v=tId35IMgPpE&t=32s> 🔗
- Final Average: 107/110
- Achieved 30/30 with a prize on Test Mining, Hardware and Software Architectures for Big Data, and Astroinformatics.

- **Courses:** data mining, programming workshop, hardware and software architectures for big data, data datamanagement and computer networks, statistical analysis of data, theory and ethics of big data, data security, uncertainty of data, test mining, data visualization, information systems and businessintelligence, astroinformatics.

Apple Developer Academy

09/2022 – 06/2023 | Naples, Italy

- Collaborated in a team to develop an audio game through brainstorming and user research sessions.
- Managed the software development lifecycle, utilizing TestFlight to conduct structured beta-testing with real users and inform iterative product refinement.
- *The app is available on the App Store* [↗](#).

Bachelor in Software Engineering, Karaj Azad University

09/2011 – 09/2016 | Karaj, Iran

Courses: Object-Oriented Programming (OOP), Data Structure and Algorithms, Database Systems Design, Operating System, Operating System Lab, Computer Architecture, Microprocessors Lab.

Publications

Modular Pipeline for Text Recognition in Early Printed Books Using Kraken and ByT5,

Yahya Momtaz et al. *Electronics* 2025, 14(15), 3083 [↗](#)

Minimizing Bleed-Through Effect in Medieval Manuscripts with Machine Learning and Robust Statistics,

Adriano Ettari, Massimo Brescia, Stefania Conte, Yahya Momtaz and Guido Russo. [↗](#)

J. Imaging 2025, 11(5), 136;

Preserving Clarity: The MAGIC project approach to ancient manuscripts,

Yahya Momtaz et al. Presented at AIUCD Conference, Verona, Italy [↗](#)

Content representation and analysis: the Magic Project and the Illuminated Dante Project integrated systems for multimedia information retrieval [↗](#)

Stefania Conte, Yahya Momtaz et al.

Artificial Intelligence in the Digital Humanities,

Journal of Artificial Intelligence Research and Innovation [↗](#)

Guido Russo, Yahya Momtaz, Sohail Maqsood

13th International Conference on Radiation, Natural Sciences, Medicine, Engineering, Technology and Ecology (RAD 2025 Conference) [↗](#)

Guido Russo, Stefania Conte, Lorenza Laccetti, Yahya Momtaz

The Magic project as a multimedia information retrieval system: content representation and analysis [↗](#)

Stefania Conte, Andrea Mazzucchi, Yahya Momtaz et al.

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/> 

Skills	
Programming Python	Machine Learning & AI PyTorch, TensorFlow, Scikit-Learn, Hugging Face Transformers, NLP (LLMs, RAG, Text Mining), Model Deployment, Fine-Tuning, Conversational AI, Chatbot Development, Agentic AI, LangChain
Data Engineering & Databases SQL, Neo4j, Data Pipelines, Data Modeling, Apache Airflow	Data Analytics & Visualization Pandas, NumPy, Matplotlib, Plotly, Tableau
Cloud & DevOps Docker, Git, GitHub Actions (CI/CD), Google Cloud Platform (GCP), AWS	Frontend React, JavaScript

Languages	
English C1	Italian B1

Certificates

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