

Mario Russo

mario@russomario.xyz | linkedin.com/in/russomarioxyz | github.com/russomario

PROFILE

AI/ML Engineer with an MSc specialization in Artificial Intelligence, hands-on experience in RAG systems, LLM evaluation, ML experimentation, and production-oriented software development. Strong Python and backend background, with practical experience in ML pipelines, embeddings, transformer models, Docker, APIs, and CI/CD.

EDUCATION

Politecnico di Milano <i>MSc in Computer Science and Engineering – Artificial Intelligence</i> • Thesis: Systematic Design and Evaluation of Retrieval-Augmented Generation for Structurally Complex Technical Documentation.	Milan, Italy 2022 – 2026
Università “Luigi Vanvitelli” <i>BSc in Computer and Electronic Engineering</i>	Caserta, Italy 2018 – 2021

AI/ML PROJECTS

RAG for Technical Documentation (MSc Thesis Project) <i>Python, RAG, Ragas, Langfuse, LLMs, Embeddings</i>	2026
- Designed and optimized a production-oriented RAG pipeline for grounded question answering over structurally complex IPC-A-610F electronics manufacturing documentation. - Evaluated layout-aware parsers, chunking strategies, embedding models, dense/sparse/hybrid retrieval, reranking, prompts, and generation models through controlled ablation studies. - Built a RAG evaluation and observability workflow using Ragas, Langfuse, LLM-as-a-Judge, and asset-aware metrics to track faithfulness, answer correctness, context quality, and figure/table reference preservation. - Identified operational trade-offs across representation quality, parsing latency, retrieval reliability, visual-reference retention, and holdout-set generalization.	
Package Theft Detection (Group Project) <i>Python, Arduino, Edge ML</i>	June 2025
- Developed a real-time, edge-oriented computer vision pipeline for package theft detection under embedded hardware constraints. - Fine-tuned a FOMO network for face/package localization and used Siamese training with knowledge distillation to deploy only the compressed face encoder on Arduino-class hardware.	
Medical QA with LLMs (Group Project) <i>Python, NLP, Transformers</i>	May 2024
- Built and compared NLP question-answering pipelines over curated medical datasets, combining indexing, clustering, embedding generation, and transformer fine-tuning. - Fine-tuned GPT-2, T5, and BERT for medical QA and compared model behavior across generative and encoder-based architectures.	
Hardware Failure Classification (Group Project) <i>Python, Scikit-learn, XGBoost, SHAP</i>	June 2024
- Built an end-to-end classification pipeline for alarm time-series data, covering preprocessing, discretization, binarization, imbalance handling, and model evaluation. - Tuned kNN, Random Forest, XGBoost, MLP, and SVM models with Hyperopt; applied SMOTE, GAN-based augmentation, and SHAP-based explainability.	

EXPERIENCE

Mirai Software <i>Software Development Supervisor</i>	Caserta, Italy Jan. 2025 – August 2025
- Analyzed client requirements, translated them into technical tasks, and coordinated delivery with developers and customer stakeholders. - Reviewed code from other developers, managed production releases, and coordinated follow-up fixes across active projects. - Monitored production applications with Grafana and Prometheus, tracking reliability issues and maintenance priorities.	
Full Stack Developer	April 2023 – December 2024
- Built Python backend services for catalog synchronization, integrating external APIs, relational databases, and scheduled data workflows. - Developed business management applications with FastAPI, PHP, Laravel, Filament, Vue, and React, using Docker-based development workflows.	
Yeet Robotics <i>Backend Developer</i>	Milan, Italy August 2022 – March 2023
- Built backend monitoring and data collection tools for retail websites, extracting structured information from external sources. - Reverse engineered APIs and profiled legacy code to improve integration reliability, performance, and maintainability.	

TECHNICAL SKILLS

AI/ML & GenAI: RAG, Ragas, LLM Evaluation, Embeddings, Transformers, Scikit-learn, TensorFlow, SHAP
Data: Pandas, NumPy, Matplotlib, JAX, SQL
Backend & APIs: Python, FastAPI, PHP, Laravel, Filament, REST APIs
MLOps & Developer Tools: Git, Docker, GitHub Actions, CI/CD, Grafana, Prometheus, Langfuse, Bash
Databases: PostgreSQL, MySQL, MongoDB, Qdrant
Frontend: React, Vue, TypeScript, JavaScript, HTML, CSS, TailwindCSS