Raffaello Fornasiere

Email: fornasiere.raffaello@gmail.com Linkedin: raffaello-fornasiere Website: rf-98.com Github: RaffaelloFornasiere

Summary

I am a software engineer with practical experience in LLM integration and AI agent development. I have built AI-powered applications in healthcare and research contexts, working across the full stack. My background combines AI/ML knowledge with hands-on development skills to deliver production-ready solutions.

Professional Experience

AI Agent Developer, Freelance for William Saunders, Alignment Science Researcher at Anthropic (May 2025 - Present)

- Building a proactive personal assistant that can autonomously initiate conversations and manage tasks.
- Implementing tool integrations (e.g. Google APIs), plus custom note-taking systems.
- Tech stack: FastAPI, React, Claude API, event-driven architecture, and various third-party APIs.

Lead Front-End Engineer, Infiniteloop (October 2022 - Present)

- Prototyped LLM integrations for healthcare applications, exploring feasibility for medical report automation (related to my thesis work),
- Designed and maintained front-end architecture using Angular, ensuring seamless integration with JHipster framework.
- Developed reusable JHipster framework extensions, including:
  - Automated JPA Specification creation system from JHipster Criteria
  - Base Repository for Spring Data with automatic bidirectional relationship synchronization via annotations
- Implemented APIs and back-end features, working across the full stack.
- Mentored two junior developers and promoted best practices to improve code quality.

Junior Full Stack Developer, Prodigys Group (July 2021 - September 2022)

- Developed and maintained both front-end and back-end components using Angular, Spring Boot, and Django.
- Optimized database performance through SQL query refactors, reducing execution times from hours to minutes.

Education

Master's in Computer Science and Engineering, Politecnico di Milano (Graduated July 2024)

- Thesis: Exploring the Potential of Lightweight LLMs for Medication and Timeline Extraction.
- Developed a web application integrating LLMs for medical information extraction, **tested in three Italian hospitals**.
- Published research: "Medical Information Extraction with Large Language Models" (ACL 2024).

Bachelor's in Electronics Engineering, Università degli Studi di Udine (Graduated July 2020)

- Thesis: Optimization of Digital Circuit Propagation Times Using Genetic Algorithms.

Technical Skills

AI/ML & LLM Integration:

- Production experience with LLM APIs (OpenAI, Anthropic Claude), and prompt engineering.
- Experience with local/lightweight models (with llamacpp) for privacy-sensitive applications.
- Agent development: Autonomous systems, tool integrations, streaming responses and conversational interfaces.
- NLP: Text extraction, information retrieval, medical language processing.
- Libraries: LangChain, llamacpp, Hugging Face Transformers.
- Academic background in neural networks (CNNs, RNNs, transformers), optimization methods.

Languages:

- Primary: TypeScript/JavaScript, Python, Java
- Also experienced with: C++, SQL.

Web Development:

- Back-end: FastAPI, NestJS, Django, Spring Boot, JHipster.
- Front-end: React, Vue.js, Angular, Tailwind CSS.
- APIs: RESTful services, third-party integrations, webhooks.

Other Tools: Docker, Git, Vector databases.

Publications

Medical Information Extraction with Large Language Models (https://aclanthology.org/2024.icnlsp-1.47/)

- Authors: **Raffaello Fornasiere**, Nicolò Brunello, Vincenzo Scotti, Mark Carman.

Academic Projects

Medical Image-Caption Matching with Transformers

– Reimplemented the CLIP model for medical applications to understand transformer architectures.

Online Learning for Dynamic Pricing in E-commerce

– Applied online learning algorithms to optimize pricing strategies in simulated environments.

Deep Learning Projects with CNNs

– Multiple image classification projects exploring neural network architectures and optimization.

Other Experiences

ML4Good AI Safety Bootcamp

– Completed EPFL's iteration of the ML4Good AI Safety and ML Bootcamp (February 2023).

Community Involvement

Volunteer developer for local community

– Built website, automated update systems, and software tools for local competition.