



Alessandro Osti

Senior Backend Engineer | AI/ML Systems

Reggio Emilia | Remote

✉ alosti78@gmail.com

☎ +39 391 708 6066

Backend & Infrastructure

Python

FastAPI

Java

MySQL

MongoDB

Docker

REST APIs

AI/ML Systems

RAG Systems

LLM Integration

Vector DBs

Semantic Search

Streaming SSE

Product & Delivery

Full Ownership

Production Systems

System Architecture

Remote Work

Profilo

Senior Engineer con 25 anni di esperienza nello sviluppo software. Ora mi concentro su prodotti basati su intelligenza artificiale.

Ho costruito e mantengo sistemi complessi in produzione: **ARIA** (assistente AI con architettura RAG, orchestrazione multi-provider LLM, ricerca semantica), **MAOTrade** (piattaforma elaborazione dati real-time in produzione dal 2012).

Mi specializzo nel portare AI da concept a prodotto finito. Competenze solide in architetture backend (Python, FastAPI, MySQL), integrazione AI/ML (RAG, vector database, LLM API) e

sviluppo di prodotti scalabili. Cerco ruoli Senior Backend Engineer o AI Engineer in product company remote-first, per costruire prodotti AI concreti con team tecnici.

Production Systems

MAOTrade - Piattaforma di trading automatico

2012 - Presente • 13+ anni in produzione • Full ownership

Piattaforma completa per elaborazione dati real-time, progettata, sviluppata e mantenuta autonomamente per oltre un decennio. Dimostra capacità di delivery end-to-end e gestione sistemi mission-critical long-running.

Stack tecnico full-stack:

- Backend: Python (FastAPI/Gunicorn), architettura containerizzata
- Database: MySQL, MongoDB per analytics e timeseries
- Frontend: Vue.js con custom charting engine (Canvas2D)
- Mobile: Android native app per monitoring real-time
- Ops: Docker, Nagios monitoring, autenticazione Google

Competenze dimostrate:

- Real-time data processing e decision systems
- External API integration (broker APIs, market data feeds)
- Production reliability: 13+ anni uptime continuo
- Full ownership: da architettura a deployment a monitoring

Key achievement: Track record comprovato di capacità di progettare, implementare e mantenere sistemi complessi in completa autonomia per periodi prolungati. Mindset production-first applicabile a qualsiasi dominio.

Python

FastAPI

MySQL

MongoDB

Vue.js

Docker

Real-time Processing

ARIA - Conversational AI System

2024 - Presente • Production-grade chatbot con RAG

Sistema conversazionale AI per query complesse, sviluppato autonomamente come progetto di apprendimento pratico. Dimostra competenze AI Engineering applicate end-to-end dalla progettazione al deployment.

Architettura e componenti:

- **Backend async:** Python FastAPI con pattern async/await, Motor (MongoDB async driver)
- **Multi-provider LLM:** Gemini 3 Flash primary (50% risparmio costi, context 1M tokens), fallback Claude Sonnet/GPT-4o per resilienza

- **RAG implementation:** Qdrant vector database, SentenceTransformers embeddings, semantic search, hybrid retrieval
- **Streaming:** SSE (Server-Sent Events) per risposta real-time
- **Conversation management:** MongoDB sessions, context optimization

Features implementate:

- Multi-intent detection con structured extraction
- Sistema a blocchi per risposte composite (grafici, tabelle, testo)
- Semantic caching per ottimizzazione costi
- Rate limiting con point-based quota
- Analytics pipeline (Fluentd, metriche token usage/latency)

Key achievement: Sistema production-ready che dimostra capacità di architettare soluzioni AI complete. Gestione autonoma intero stack (LLM orchestration, vector search, streaming, monitoring) con attenzione a costi e scalability.

RAG Systems

Multi-LLM Architecture

FastAPI Async

Qdrant

Vector Embeddings

SSE Streaming

MongoDB

Esperienza Professionale

○ Senior Backend Developer

[Lynx S.p.A.](#)

Apr 2018 - Presente • Remote

Sviluppo backend per applicazioni enterprise. Esperienza su architetture complesse con requisiti di affidabilità e performance.

- Backend architectures: REST APIs, microservices, system integration
- Database optimization: Oracle/PostgreSQL query tuning
- Production support: troubleshooting, incident resolution

Java

Spring Boot

Oracle

PostgreSQL

REST APIs

○ Software Engineer

[Vari clienti enterprise](#)

1999 - 2018

19 anni di sviluppo sistemi production per sanità, banking ed enterprise. Progetti principali:

- Sistemi gestionali sanitari: Sviluppo full-stack con architetture Oracle complesse, ottimizzazione PL/SQL, app Android con integrazione NFC/GPS per rilevazione presenze e raccolta dati
- Sviluppo mobile: Applicazioni Android native con architettura offline-first, barcode scanning, geolocalizzazione, connettività Bluetooth/NFC per dispositivi IoT

- Integrazione enterprise: Co-fondatore società di consulenza specializzata in system integration—sviluppo middleware per connessione sistemi ERP (Navision), tool di migrazione email (Lotus Notes verso SharePoint/Google Workspace), soluzioni mobile multi-platform
- Sistemi telefonici: Sviluppo sistemi IVR e driver C++ per infrastrutture telecom, applicazioni real-time per gestione chiamate in contact center

Competenze trasferibili: Architetture backend scalabili, debugging production sotto pressione, delivery autonomo di progetti, collaborazione team distribuiti, ownership completo di sistemi complessi da progettazione a manutenzione.

Tecnologie: Java, C++, Oracle, PL/SQL, Android, IBM enterprise stack, integrazione ERP.

Formazione

AI/ML Systems - Self-directed Learning (2024-2025)

- LLM fundamentals: transformer architectures, prompt engineering, context management
- RAG systems: vector databases, embedding models, retrieval strategies
- Production deployment: FastAPI async, streaming, error handling, monitoring
- Implementation pratica: ARIA chatbot (12 settimane structured learning)

MBA in Finance | Ateneo Borsari

Punti di Forza

- **End-to-end delivery:** Capacità comprovata di portare progetti dalla progettazione al deployment in autonomia
- **Production mindset:** Focus su reliability, monitoring, error handling, cost optimization
- **Full-stack capability:** Comfort con intero stack quando necessario (backend, database, frontend, ops)
- **Fast learner:** Transizione AI in 12 mesi con progetto production-ready
- **Long-term reliability:** 25 anni esperienza con sistemi mission-critical

Disponibilità: Remote preferred. Hybrid accettabile (Milano/Bologna 2-3 giorni/settimana). Base Reggio Emilia.

Autorizzo il trattamento dei miei dati personali ai sensi del Decreto Legislativo 30 giugno 2003, n. 196 e dell'art. 13 del Regolamento UE 2016/679 (GDPR) ai soli fini della ricerca e selezione del personale.

