

RENZO ALBERTO LENES CHUQUICAÑA

AI Engineer | GenAI · LLMs · LangChain · Python · Cloud (AWS)

Lima, Peru • +51 904 890 457 • lenesrenzoalberto@gmail.com

linkedin.com/in/renzo-alberto-lenés • github.com/RenzoLenes • portfolio-fullstack-topaz.vercel.app

PROFESSIONAL SUMMARY

AI Engineer with 3+ years of experience developing and deploying Generative AI (GenAI) solutions in production. Specialized in Large Language Models (LLMs) with OpenAI, advanced prompting, RAG architectures and agent orchestration with LangChain and equivalent frameworks. Strong command of Python for building APIs, pipelines and LLM integrations. Experienced in designing scalable, secure architectures with Infrastructure as Code (IaC) and cloud deployment on AWS. Focused on delivering production-ready, governed and maintainable AI solutions while collaborating with cross-functional teams.

KEY SKILLS

GenAI & LLMs: OpenAI, Claude, Gemini, advanced prompting, RAG, embeddings, semantic search

AI Frameworks: LangChain, Mastra (agent orchestration), function calling, multi-agent systems

Python: APIs (FastAPI), pipelines, MCPs, LLM integrations, pandas/numpy

IaC & Cloud: AWS CloudFormation (IaC), Terraform (transferable base), AWS (EC2, S3, Lambda), GCP

Architecture: Scalable, secure and governed system design for production AI

Tools: Claude Code, GitHub Copilot, Docker, Git, n8n

Languages: Spanish (native), English B1 intermediate

PROFESSIONAL EXPERIENCE

VentlA | *Head of Fullstack Development & AI Engineer* | Jan 2026 – Present · Remote

- Designed, developed and deployed GenAI solutions in production: intelligent agents powered by LLMs (OpenAI, Gemini), RAG architectures and automation pipelines that transformed real client business operations.
- Built Python components — FastAPI APIs, data pipelines and LLM integrations — orchestrating agents with LangChain and Mastra for complex multi-step workflows.
- Designed and tested advanced prompts iteratively to achieve precise, relevant and controlled outputs, reducing hallucinations and ensuring predictable behavior in production.
- Designed scalable and secure architectures for AI solutions, applying Infrastructure as Code with AWS CloudFormation for consistent and reproducible deployments.
- Collaborated with product and operations teams to integrate AI solutions into production environments, documenting business rules and ensuring maintenance best practices.

VentlA | *AI Engineer* | Nov 2025 – Jan 2026 · Remote

- Built RAG pipelines in Python with chunking, embeddings and semantic reranking, integrating LLMs over proprietary knowledge bases for reliable, context-aware responses.
- Experimented with agent orchestration frameworks (LangChain, Mastra), evaluating approaches before committing them to production.

Freelancer – AI & Software Engineer | *AI Engineer / ML Practitioner* | 2024 – Present

- Developed Servinetwork, a Python-based RAG and semantic search platform that autonomously resolves ~80% of queries, with explicit validation of AI-generated outputs.

- Built a sales forecasting model for a textile retail company using XGBoost and Random Forest (pandas, numpy, scikit-learn), deployed to production by the client.
- Integrated LLM APIs (OpenAI, Claude) into production applications, designing scalable architectures deployed on AWS with a focus on security and data governance.

ENTEL | *Data Analysis Intern* | May 2025 – Oct 2025

- Developed data pipelines in Python (pandas, numpy) over high-volume financial datasets, applying statistical transformations and process automation.

EDUCATION

B.S. in Information Systems Engineering

Universidad Peruana de Ciencias Aplicadas (UPC) • 2021 – 2026

FEATURED AI PROJECTS

Servinetwork – Python RAG System (Production)

Full RAG pipeline in Python: chunking, embeddings, vector search and reranking. LLM orchestration with LangChain. OpenAI API. Advanced prompting for hallucination control. Autonomously resolves ~80% of queries with output validation.

VentIA – GenAI Agents for Automation

Intelligent agents with LLMs and function calling orchestrated with LangChain and Mastra. Python APIs (FastAPI), external system integrations and AWS deployment with Infrastructure as Code (CloudFormation). 60% reduction in manual handling time.

Sales Forecasting Model — Textile Retail (private project, 2025)

Forecasting with XGBoost and Random Forest. Full Python pipeline (pandas, numpy, scikit-learn), seasonal feature engineering and time-based validation. Deployed to production for inventory planning.

TRAINING & CERTIFICATIONS

- LLM Engineering: Master AI, LLMs, RAG, QLoRA & Agents – Ed Donner (Udemy) (2025)
- AWS Certified Cloud Practitioner – Amazon Web Services (2025)
- Node.js: Zero to Expert (Clean Architecture, DDD, PostgreSQL) – F. Herrera (2024)
- Training Camp: Advanced Data Structures – UPC (2023)

TECH STACK

Python · LangChain · Mastra · OpenAI API · Claude · Gemini API · LLMs · RAG · Embeddings · Semantic search · Advanced prompting · Function calling · Agents · MCPs · FastAPI · pandas · numpy · scikit-learn · XGBoost · AWS (EC2, S3, Lambda) · CloudFormation (IaC) · Terraform · GCP · Docker · Git · Claude Code · GitHub Copilot · n8n