

Carlos Perez Molero

Forward Deployed AI Engineer & Tech Lead

perez.moleroc@gmail.com | +54 9 11 2251 0584 | Buenos Aires, Argentina |

linkedin.com/in/carlos-perez-molero | github.com/engcarlosperezmolero | charlymolero.com

PROFESSIONAL PROFILE

Forward Deployed AI Engineer and Tech Lead with 7+ years of experience building production-grade software and data systems. I embed with client teams, run discovery, own the solution architecture end to end, and lead delivery through deployment and handoff. Recently focused on AI agents, LLM-powered applications, RAG systems, LLM evaluation, and cloud deployments, integrating OpenAI, Claude, and Gemini into real user workflows. Background in Chemical Engineering, valuable for industrial, energy, and process-oriented AI applications.

EXPERIENCE

HUMAI Institute

02/2025 - Present

Forward Deployed AI Engineer & Tech Lead · Contractor

- Single point of contact for clients: requirements gathering, weekly meetings, demos, handoff, and interviewing engineers for staff augmentation
- Coordinated teams of external AI experts to produce cutting-edge AI courses: interviewed and negotiated with instructors, wrote scripts and teaching guides, and ran QA on every recorded class
- Built full-stack applications with FastAPI, PostgreSQL (Cloud SQL), GCP (Cloud Run, GCS), GitHub Actions, and Docker
- Implemented secure RBAC, cookie-based authentication, and real-time scraping with Playwright
- Designed ETL pipelines with VertexAI (LLM-based and RAG) for data processing and analysis
- Developed React frontend with SSE, Kotlin companion app, and backend using FastAPI with CI/CD
- Taught 100+ classes on Python, automation, web scraping, databases, CI/CD, and Git

AI, Python, FastAPI, React, Next.js, TypeScript, PostgreSQL, GCP, AWS, Docker, Playwright, VertexAI, Langfuse, Kotlin, GitHub Actions

Conversational AI Interview Platform

LLM-powered conversational interview platform for qualitative research: an AI agent conducts in-depth interviews by text or voice, decides when to probe deeper, and consolidates findings. Ran discovery, biweekly demos, and handoff directly with the research team, and selected models through LLM-as-judge evals on real interviews. Built the complete system end-to-end on AWS.

AI, Next.js, TypeScript, Claude, AWS Bedrock, PostgreSQL, AWS

Parental Monitoring Platform

LLM-powered parental monitoring application that scrapes messaging platforms using whatsmeow and processes conversations through multiple LLM layers to generate emotional state metrics and alerts. Worked directly with the founders from beta to funding. Built the complete system end-to-end: backend, frontend, scraper, LLM analysis service, alerting, and observability with Langfuse on GCP.

AI, Go, whatsmeow, OpenAI, Claude, Langfuse, GCP, React, FastAPI

Music Industry Management App

AI-driven application for music managers that uses LLMs to generate and execute artist launch plans. Worked directly with the founder, led a team of 3 engineers, and owned model and architecture decisions: migrated Django to FastAPI with zero downtime and cut LLM cost per plan by 88%.

AI, React, FastAPI, Django, OpenAI, PostgreSQL

Government Process Automation

Automated remote processing system for government registry procedures, replacing manual filing by the client's operations team. Gathered requirements with operators and the PM, directed the client's technician, and put into production a self-healing agent pipeline that recovers lost sessions, resumes from the failed step, and never files twice. Built with Playwright, backend API, dashboard, and batch processing.

AI, Playwright, FastAPI, React, PostgreSQL

Banking Data Migration

Senior Data Engineer migrating a daily reconciliation process from on-prem IBM AS400 to the cloud using AWS, Databricks, Airflow, Spark, Kafka, and Kubernetes. Led the 10-person team's AI adoption: Claude Code skills, hooks as guardrails, and secure MCP servers to on-prem and observability tools. Optimized ETL pipelines and DAGs, reducing processing times by over 50%.

AI, AWS, Databricks, Airflow, Spark, Kafka, Kubernetes, Python

Techunting
12/2023 - 02/2025

Sr. Full-stack Python Developer · Contractor

- Developed the AWS-based backend for a super yacht brokerage system using ECS and FastAPI, designing a PostgreSQL database with 20 tables and building 34 API endpoints
- Implemented an OpenSearch-powered search engine and managed images with S3 and Cloudinary, with Postgres triggers for logging and webhooks for internal and client use
- Built complex ETL pipelines to ingest and unify data from client MySQL data lakes into PostgreSQL, ensuring seamless data integration
- Designed PDF generation workflows with Jinja2 templates and macros, and built LLM-powered PDF data extractors using OpenAI API to structure unstructured document data
- Created comprehensive Pytest test suites for unit and integration testing with 80%+ coverage
- Led a small team to deliver features, manage releases, and implement streamlined CI/CD workflows with GitHub Actions and Docker
- Supported UI development with React and TypeScript for full-stack collaboration

AI, Python, FastAPI, React, TypeScript, PostgreSQL, OpenSearch, AWS, Docker, GitHub Actions, Jinja2, Pytest, S3, Cloudinary, OpenAI

Ernst & Young (EY)
01/2022 - 12/2023

Ssr. Backend Python Developer · Contractor

- Architected and maintained 9 microservices with FastAPI and MongoDB
- Implemented continuous deployment using Kubernetes through Azure DevOps
- Executed data migrations with PySpark, Pandas, and Pymongo on Azure Databricks
- Developed backend microservices with ASP.NET (C#)

Python, FastAPI, MongoDB, Kubernetes, Azure, PySpark, Pandas, Databricks, ASP.NET, C#

Equifax
03/2021 - 12/2021

Data Solutions Developer · Full-time

- Built data applications using Python, HTML, CSS, JavaScript, and SQL
- Managed a 15-member development team on GitHub
- Migrated ETL processes from local server to GCP

Python, JavaScript, SQL, GCP, HTML, CSS

Nerd Interactive
07/2020 - 02/2021

Data Science Analyst · Full-time

- Handled Big Data through SQL in BigQuery from Google Cloud Platform (GCP)
- Performed data analysis in Python (Pandas) to create KPIs
- Developed REST APIs with Django

Python, Pandas, SQL, BigQuery, GCP, Django

TECHNICAL SKILLS

AI & LLM: OpenAI API, Claude API, Gemini / VertexAI, AWS Bedrock, LangChain, LangGraph, Langfuse, RAG Systems, Prompt Engineering

Backend: Python, FastAPI, Django, Node.js, TypeScript, ASP.NET / C#

Frontend: React, TypeScript, HTML & CSS, Tailwind CSS

Data Engineering: Airflow, Spark / PySpark, Kafka, Databricks, BigQuery, ETL Pipelines

Cloud & DevOps: AWS, GCP, Azure, Docker, Kubernetes, GitHub Actions CI/CD

Databases: PostgreSQL, MongoDB, Redis, Opensearch

Testing & Monitoring: Pytest / UnitTest, Playwright, Datadog, Grafana, CloudWatch

PROJECTS & CLASSES

Trampoline Gymnastics AI App - RAG-based
2025 · Buenos Aires, AR

Full-stack app for trampoline gymnastics with a difficulty calculator and scoring, plus an AI assistant that answers questions about the FIG Code of Points using RAG over the rulebooks and tool calls to the difficulty engine

AI, React, TypeScript, FastAPI, Gemini, pgvector, Langfuse

https://tra-app.clever-club.com/bot

Data Analysis Class
2022 · Buenos Aires, AR

Data analysis course dictated by me at HUMAI

Python, Pandas

https://www.youtube.com/watch?v=V2t4brIm6xM

FastAPI Workshop
2025 · Buenos Aires, AR

FastAPI workshop: ideal for students, professionals, and researchers who want to learn with a real-world example

Python, FastAPI

https://youtu.be/EjY_KGjovIs

Tuclases.com Web Scraper 2022 · Buenos Aires, AR

Automated web scraping tool for educational platform data extraction

Python, Automation

https://github.com/engcarlosperezmolero/tus_clases_automatic_response_scraper

Time of Fly Measurement ESP32 Device 2022 · Buenos Aires, AR

IoT device for time-of-flight measurement using ESP32 microcontroller with web interface

C++, JavaScript, HTML, CSS

https://www.linkedin.com/feed/update/urn:li:activity:6982465610228088832/

Ser Dev en Tiempos de IA 2025 · Buenos Aires, AR

Live stream about the real advantages of being a developer in the age of AI, covering topics shaping the present of software development

AI, Development

https://www.youtube.com/live/t8o2N41fo1I

EDUCATION

B.S. Chemical Engineering (Ingeniero Químico, 5-year degree) 02/2013 - 06/2019

Universidad del Zulia

Grade average: 18.66 / 20 (GPA equivalent: 9.33 / 10) | Graduated with honors

Internship at Indesca working with a PVC production pilot reactor. Thesis developed using ASPEN and HYSYS process simulation software.

LANGUAGES

Spanish: Native

English: Advanced | Professional