

Lucas Romero Mercado

AI/ML Engineer & Full-Stack Developer

Bogotá, Colombia | Open to Remote | +57 304 573 9227 | lucasromero2605.aot@gmail.com
linkedin.com/in/lucasromerom | github.com/LucasRomero26 | lucasromero.site

SUMMARY

AI/ML Engineer and Full-Stack Developer who takes machine learning from notebook to production: deep learning for computer vision and Large Language Model (LLM) applications, plus scalable Python and TypeScript web systems. Experience leading a software & ML research team, fine-tuning and evaluating LLMs, and shipping end-to-end products with measurable results—91.7% segmentation mAP, ~47 ms in-browser inference, and secure multi-tenant AI assistants. Comfortable across the full stack: dataset curation, model training and optimization, the API that serves it, and the web app that puts it in users' hands.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL, C/C++ (embedded), Bash

AI & LLM: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), tool calling / AI agents, prompt engineering, LLM fine-tuning & evaluation, Model Context Protocol (MCP), conversational AI, NVIDIA NIM (Llama 3.1), OpenAI API, LangChain, Hugging Face

Machine Learning & Computer Vision: PyTorch, Ultralytics YOLO (detection & segmentation), OpenCV, scikit-learn, statsmodels (SARIMAX / SARIMA), time-series forecasting, model export & quantization (LiteRT / TFLite, ONNX), edge inference (WebGPU)

Frontend: React, Next.js, Vite, TypeScript, Tailwind CSS, shadcn/ui

Backend: Node.js, Fastify, Express, NestJS, FastAPI, REST APIs, PostgreSQL, MongoDB, Supabase, Redis, JWT authentication, OAuth 2.0

MLOps, DevOps & Cloud: Docker, GitHub Actions (CI/CD), Vercel, Render, AWS (Cloud Foundations), Prometheus, Sentry, observability & monitoring, load testing (k6)

EXPERIENCE

Software Development Lead — GIRSI Research Group, Universidad del Norte

Feb 2025 – Jun 2026

Barranquilla, Colombia

- Lead the software and machine-learning workstream for a maritime-logistics line, building deep-learning computer-vision models for automated shipping-container code recognition (OCR).
- Designed and delivered end-to-end software and embedded firmware for the "Interactive Forests" IoT project, spanning device firmware through the application layer.
- Mentor student researchers in machine-learning fundamentals, software architecture, and project-based R&D execution.

AI Coding & Evaluation Specialist (Latam Coder) — Outlier

Jul 2025 – May 2026

Remote

- Contributed to training, fine-tuning, and evaluation of advanced LLMs through full-stack application development.
- Integrated Model Context Protocol (MCP) and conversational-AI architectures to improve context handling and response accuracy.
- Wrote test suites, triaged and resolved software bugs, and curated multimodal datasets via image annotation and technical transcription to ensure model-training data quality.

PROJECTS

KillaAssistant — Multi-Tenant AI Assistant

Next.js, Fastify, TypeScript, Supabase, Docker

killassistant.vercel.app | github.com/LucasRomero26/killassistant

- Built a multi-tenant AI assistant over Telegram and WhatsApp that orchestrates up to 5 tool-calling rounds against NVIDIA NIM (Llama 3.1 70B) for Google Calendar and Drive automation.
- Secured per-user LLM keys and OAuth tokens with AES-256-GCM encryption, Supabase row-level security, JWT auth, and a prompt-injection sanitizer; added Groq Whisper voice transcription.
- Shipped with Docker and GitHub Actions CI/CD (typecheck, lint, tests) and automated deployment with health checks.

EpiGuard — Dengue Early-Warning ML Platform

Python, statsmodels, Next.js, Supabase

epiguardapp.vercel.app | github.com/LucasRomero26/epiguard

- Benchmarked six model families (SARIMA, SARIMAX, MLP, ElasticNet, Random Forest, XGBoost) to forecast dengue risk 4 weeks ahead across Colombia's 32 departments, achieving a weighted RMSE of 25.28.
- Deployed interpretable SARIMAX to attribute each forecast to specific climate variables; integrated SIVIGILA surveillance, NASA POWER meteorology, and NOAA ONI data.
- Served ML inference on Hugging Face Spaces behind an interactive Next.js choropleth dashboard, with automated daily predictions via GitHub Actions.

SugarcaneLeaf — Edge ML Disease Segmentation

PyTorch, YOLO, WebGPU, FastAPI, React

sugarcaneleaf.vercel.app | github.com/LucasRomero26/sugarcaneleaf

- Fine-tuned a YOLO segmentation model on 6,000+ field images across 5 disease classes, reaching 91.7% mAP@0.5.
- Exported to LiteRT / ONNX and served entirely in-browser via WebGPU—no server roundtrip, ~47 ms p99 client-side inference with automatic WASM fallback.
- Built a FastAPI backend with a Prometheus-instrumented latency ring buffer and a live p50/p99 MLOps dashboard; validated PyTorch-to-LiteRT export parity (96%) and load-tested with k6.

Granger — Full-Stack Project Management App

MERN, TypeScript, Redis, Docker

grangerproject.vercel.app | github.com/LucasRomero26/granger

- Built a MERN (MongoDB, Express, React, Node.js) app with end-to-end TypeScript, Kanban boards, role-based authorization, and Google/GitHub OAuth login.
- Implemented JWT refresh-token rotation with reuse detection and chain revocation, httpOnly/SameSite cookies, and four Redis-backed rate-limit buckets.
- Covered by Playwright end-to-end and Vitest tests (mongodb-memory-server); instrumented with Sentry and Better Stack across a Vercel/Render deployment.

EDUCATION

Universidad del Norte

Barranquilla, Colombia

- B.Sc. in Electrical Engineering
- B.Sc. in Electronics Engineering

CERTIFICATIONS

- Artificial Intelligence Technology and Applications — Huawei (2025)
- AWS Academy Cloud Foundations — Amazon Web Services (2025)
- Full-Stack Development with Node.js, React, TypeScript, NestJS & Next.js — Udemy (2025)
- React and TypeScript: Building 10+ Projects — Udemy (2025)
- Node.js Web Development Bootcamp (MVC & REST APIs) — Udemy (2025)
- Data Analysis & Visualization + AI — Udemy (2025)
- Python University: From Zero to Expert (114+ hours) — Udemy (2026)

LANGUAGES

Spanish (Native) • English (Professional Working Proficiency) • Italian (Professional Working Proficiency)