

Juan Javier Ramírez Magaña

Applied AI · Automation · Local LLM Infrastructure

Mexico City, Mexico · Open to remote roles · +52 55 7885 7991 · jjavierrm@gmail.com · github.com/Kirajav

PROFESSIONAL SUMMARY

Engineer with 15+ years of hands-on technical experience, now building private AI infrastructure: local models on Ollama, a semantic router that chooses local or cloud models based on how confidential each request is, MCP servers that let an agent manage Proxmox and DNS, and Whisper transcription. Strong foundation in networking, servers, Python and automation, plus experience scoping AI solutions (RAG, chatbots, transcription) for SMBs.

TECHNICAL SKILLS

Applied AI: Ollama and local models (Qwen, Llama, DeepSeek-Coder); MCP servers (FastMCP); local/cloud LLM routing (AWS Bedrock, Gemini); Whisper; RAG and chatbot solution design

Systems & infrastructure: Windows and Windows Server, Linux; Proxmox VE (VMs and LXC); Docker and Docker Compose; backups and monitoring; UPS

Programming: Python, C/C++, JavaScript/Node.js, PowerShell, SQL, HTML/CSS, PHP; React and Vue; .NET and VBA; Git/GitHub

Data & BI: SQL (PostgreSQL, MySQL), Power BI (DAX, Power Query), Python (Pandas), advanced Excel; APIs (Mercado Libre) and web scraping

Networking: LAN/WLAN, routers, switches and firewalls; GPON fiber/ONT; DNS (AdGuard Home); PoE IP video surveillance

Embedded & IoT: ESP32 / ESP32-S3 / ESP32-C3, Arduino, Raspberry Pi; C/C++ with PlatformIO and FreeRTOS; MQTT, Home Assistant, ESP-NOW, OTA; LVGL; I2C, SPI, UART; ISP programming

PROFESSIONAL EXPERIENCE

Senior IT & IoT Consultant · Self-employed (freelance)

2024 – Present

Mexico City · remote and on-site

- Designed and run a Proxmox homelab (VMs, LXC and Docker) hosting a private AI stack: local models on Ollama, a semantic router that keeps confidential prompts on local models and sends the rest to cloud LLMs, and an MCP server that lets a Telegram agent manage Proxmox and AdGuard DNS.
- Automate workflows with Python and PowerShell: tablet-based PC control (per-app audio, DDC/CI monitor control, window management, keyboard and mouse) and auto-generated technical reports with diagrams.
- Scope digital-platform (website, email, domain) and AI service proposals for SMBs (RAG assistant, WhatsApp chatbot, transcription), including architecture, cost model and rollout plan.
- Provide remote and on-site IT support to accounting and architecture firms: networks, workstations, UPS units and Windows servers (setup, backups and monitoring).
- Design PoE IP video-surveillance installations: camera selection, cabling plan, connection diagram and a visual mock-up of each camera's placement for the client.

Infrastructure Specialist · Totalplay

2022–2023

Mexico City

- Supervised GPON fiber-optic data services; provisioned and configured ONTs, routers and TV equipment.
- Configured network security (firewalls and switches) and installed wired and wireless networks for customers.

Data Analyst & Junior Developer · Vetco Gray (GE)

2021

Mexico City

- Built real-time KPI dashboards with .NET, Excel/VBA and SQL database connections.

Independent IT Technician & Consultant · Self-employed

2011–2020

Mexico City · on-site and remote

- Repaired and maintained laptops, PCs, UPS units and printers for homes and businesses, with written diagnostics and quotes.

Technical Support Consultant · Iusacell

2010–2011

Mexico City

- Provided specialized technical support to mobile users and provisioned data services remotely.

Programmer Analyst · RAM Software y Soluciones

2004–2006

Mexico City

- Developed C and C++ systems, including data validation, testing and technical documentation.

SELECTED PROJECTS

Private AI stack on a homelab

2026

Semantic request router (local vs. cloud, fail-closed for confidential data) with a rules dashboard, an MCP server for Proxmox and AdGuard, and a Whisper transcription server. Stack: Python, FastMCP, Docker, Ollama, Proxmox.

Tablet control surface for Windows

2026

Python bridge that turns a tablet into a control surface: VoiceMeeter mixing, per-app volume and audio routing, DDC/CI monitor control, window manager, keyboard and mouse. Stack: Python, OSC/UDP, Win32/COM, Node.js.

JamJum JP-1 Studio — MIDI controller reverse engineering

2026

Decoded the device's SysEx protocol (pads, knobs, buttons, RGB color) and its checksums, validated 384/384 against real captures, and built a web editor with RGB effects and automated tests. Stack: Python, JavaScript, Web MIDI, Wireshark/USBPcap, Node.js.

Diagnostic report generator

2026

Internal tool that assembles diagnostic and quote reports with SVG diagrams of the device (memory, storage, battery and disk health). Stack: Python, SVG, Word.

EDUCATION

Computer Engineering

2005–2008

School of Engineering, National Autonomous University of Mexico (UNAM)

Certifications in Progress: Cisco CCNA certification, in preparation

Languages: Spanish (native) · English (advanced, C1) · Japanese (basic)