

RAMIRO CARNICER SOUBLE

Buenos Aires, Argentina | +54 11 5782 3233 | ramirocarnicersouble8@gmail.com | ram4.dev | linkedin.com/in/ramirocarnicersouble | github.com/ram4-dev | x.com/ram4_dev

Generative AI Engineer / Applied AI Product Builder | LLM Agents | RAG | Backend Systems | SQL & Python

PROFILE

Applied AI and backend engineer focused on taking generative AI from prototype to production adoption. Combines production backend experience at Mercado Libre with hands-on LLM agent development as Co-founder/Product Engineer of Khora. Builds practical AI systems: agents, RAG workflows, automation layers, and backend integrations that connect with real business processes while respecting guardrails, observability, security, and cost constraints.

APPLIED AI ENGINEER FOCUS

AI adoption & business impact	Identifies high-value automation opportunities, scopes MVPs, defines success metrics, and translates business problems into practical AI workflows.
LLM agents & workflows	Builds flows with planning, tool use, context handling, retrieval, evaluator/reviewer loops, and human-in-the-loop escalation.
RAG & knowledge access	Designs retrieval patterns over company context using vector search, structured filters, prompt/context management, and source-grounded responses.
Systems integration	Connects LLM solutions with APIs, internal tools, SaaS platforms, databases, webhooks, scheduled jobs, and backend services.
Responsible production mindset	Focuses on reliability, monitoring, failure handling, security boundaries, cost awareness, auditability, and reusable documentation.

CORE CAPABILITIES

Generative AI: LLM applications, agents, LangGraph, tool-calling, RAG, pgvector/vector databases, prompt systems, evaluator loops, AI workflow design

Data & backend: SQL, PostgreSQL, Python, Java/Spring Boot, Node.js, Django, REST APIs, microservices, background jobs, integrations

Product & production: use-case discovery, process mapping, stakeholder translation, MVP scoping, observability, safe AI usage, documentation

PROFESSIONAL EXPERIENCE

Co-founder / Product Engineer | Khora - AI-native Organizational Intelligence

Buenos Aires, Argentina | 2024 - Present | khora.ar

- Building an AI-native platform that automates internal workflows, centralizes organizational knowledge, and produces actionable reports for decision-making.
- Design agentic workflows where AI retrieves company context, answers process questions, invokes tools, and escalates human-in-the-loop decisions when risk or uncertainty requires it.
- Work on RAG/context architecture, backend integrations, data model decisions, LLM orchestration strategy, product UX, and demo-driven iteration.

- Own discovery, partner conversations, user feedback, and MVP scoping under uncertainty, balancing technical feasibility with business value and adoption.

Backend Software Engineer | Mercado Libre

Buenos Aires, Argentina | Dec 2025 - Present

- Engineer backend services inside Latin America's largest e-commerce and fintech technology ecosystem, where reliability, latency, scale, and maintainability are core constraints.
- Work on scalable APIs and distributed backend systems, applying production-grade standards across performance, observability, debugging, and failure handling.
- Use AI tools to accelerate engineering workflows while keeping architectural judgment, code quality, security, and production safety as controlling constraints.

Full Stack Developer - Contract for Arcor Group | AYI Group

Buenos Aires, Argentina | Feb 2025 - Dec 2025

- Implemented reactive microservices with Java Spring Boot for marketing automation workflows processing 1,000+ messages/day.
- Built recurrent background jobs and backend services using Node.js/Django, integrating persistence, automation logic, and operational flows across backend/database/application layers.

SELECTED AI / AUTOMATION BUILDS

Designer / Builder | SDD Custom Kit - Spec-driven Multi-agent Development

Open-source / GitHub | In Progress

- Built a portable workflow using structured specs and specialized AI agents to move from requirements to implementation, with planner/developer/reviewer loops and evaluation checkpoints.

Creator / Architect | Infer - Community-powered LLM Inference

infer.ram4.dev | In Progress

- Designed an OpenAI-compatible inference gateway concept for specialized community-hosted models, focused on lower cost, provider routing, and no-vendor-lock-in access.

EDUCATION

- University of Buenos Aires (UBA) - Bachelor's in Informatic Engineering, 2nd year. Focus: software engineering, algorithms, and systems foundations.
- Roberto Rocca Technical School - Electronic Technician. Technical degree with C, C++, Python, hardware logic, electronics, and first-principles engineering foundation.

LANGUAGES

- Spanish: Native | English: B2 / Professional working proficiency