

# Carlos Machicado

Desarrollador Backend & IA Aplicada | Java · LLMs on-premise · RAG · Privacy-by-Design  
(+51) 987 099 443 · [carlosmachicad@gmail.com](mailto:carlosmachicad@gmail.com) · [linkedin.com/in/c-machicado123](https://linkedin.com/in/c-machicado123) · [github.com/cma208](https://github.com/cma208)

## PERFIL PROFESIONAL

Desarrollador Backend con más de 5 años de experiencia y un fuerte enfoque en Inteligencia Artificial aplicada. Diseño y construyo agentes de IA y soluciones RAG con modelos de lenguaje (LLMs) ejecutados 100% en local (on-premise) y modelos open source, sobre una base backend sólida en Java 17/21 y Spring. Responsable de arquitectura y desarrollo end-to-end, con foco en sistemas de IA privados y soberanos —los datos nunca salen a terceros— mantenibles y seguros.

## EXPERIENCIA PROFESIONAL

### Universidad Ricardo Palma

Lima, Perú

#### Desarrollador Backend · Área de Gestión de Proyectos

Feb 2023 – Actualidad

- Diseñé y construí end-to-end un agente conversacional RAG 100% on-premise (Java 21, Spring Boot reactivo, Spring AI): recuperación híbrida (vector + BM25 en español con fusión RRF) y reranking sobre PostgreSQL + pgvector, citación inline obligatoria y guardrails anti-alucinación y anti-jailbreak en capas; ~239 clases con suite de pruebas (Testcontainers, k6).
- Serví los LLMs localmente en una sola GPU (vLLM/Ollama, sin APIs comerciales) sosteniendo 30 sesiones concurrentes a p50 7 s, con autorización por rol previa al retrieval, auditoría inviolable (hash-chain SHA-256) y cumplimiento de protección de datos (Ley 29733 / GDPR). Validado en pruebas de carga e iniciando pruebas con usuarios (pre-piloto).
- Contribuí, en el equipo de desarrollo, a varios módulos del nuevo sistema académico “Yachayhuasi” (matrícula, carga lectiva y no lectiva, cuadro de necesidades, dashboards, reportes), en backend (Java 17/21, Spring) y luego frontend; migré código legacy (Visual Basic, Java, PHP) a Java moderno y Python.
- Participé en la renovación del portal web institucional, reconstruí el módulo de Libro de Reclamaciones (aplicación consolidada y mantenible) y desarrollé reportes automatizados en PDF y Excel con Java (Apache POI).

### Ganemo

Lima, Perú (Remoto)

#### Desarrollador Junior

Feb 2022 – Nov 2022

- Desarrollé y mantuve plugins para los productos de software de la cartera de clientes, y automatice procesos administrativos conectando vistas y widgets con la base de datos.

### Pactomática

Chile (Remoto)

#### Trainee de Desarrollo

Mar 2021 – Nov 2021

- Automatice procesos de contratación, RR.HH. y gestión de contratos (“despapelización”), detectando y corrigiendo errores en los flujos de automatización.

## HABILIDADES TÉCNICAS

**IA & LLMs:** Agentes de IA, RAG, LLMs on-premise (Ollama, llama.cpp, vLLM), Spring AI, LangChain4j, modelos open source (Llama, Mistral, Qwen, DeepSeek), bases vectoriales (pgvector, Qdrant, Chroma), NLP, Machine Learning

**Backend & APIs:** Java 17/21, Spring / Spring Boot, Python, Django, API REST, Apache POI

**Bases de Datos & Análisis:** SQL Server, Oracle, PostgreSQL, MySQL, Pandas, NumPy

**DevOps & Herramientas:** Docker, Docker Compose, CI/CD (GitHub Actions), Git, GitHub, Tomcat

**Frontend:** React, JavaScript, HTML5, CSS3

**Otros lenguajes:** PHP, Visual Basic

**Idiomas:** Español (Nativo) · Inglés (C1)

## EDUCACIÓN

### Maestría en Ciencia de Datos (Egresado)

Mar 2023 – Jun 2025

Universidad Ricardo Palma

### Bachiller en Ingeniería Mecatrónica

Mar 2014 – Ago 2020

Universidad Ricardo Palma

## PROYECTOS DESTACADOS

- **NEXUS CORE** — case study del agente RAG on-premise: [cma208.github.io/nexus-core-study-case](https://cma208.github.io/nexus-core-study-case)
- Deep Learning / ML: predicción de fuga de clientes con redes neuronales convolucionales y clasificación de tumores para detección de cáncer de mama.