

Jonathan Joan Avila Huamolle

Senior AI Engineer | Backend & Fullstack | Generative AI (RAG) · Python · AWS

Tacna, Perú · jonathan.joan.avila@gmail.com · +51 940 202 780 · [LinkedIn](#)

PROFESSIONAL SUMMARY

Senior AI Engineer with a strong backend and fullstack foundation — 8+ years building scalable systems and 3+ years taking Generative AI from prototype to production. I design RAG architectures, LLM agents and NLP document-processing pipelines in Python, with a constant focus on reliability, accuracy and cost at scale (model routing, caching, evaluation, human-in-the-loop). Deep expertise in AWS serverless and event-driven architecture (Lambda, API Gateway, EventBridge, SQS/SNS, Bedrock, CDK), plus fullstack range with React/TypeScript and hands-on Azure. Proven delivery in fintech and document-heavy, regulated domains.

Core strengths:

- **Generative AI in production:** RAG systems, LLM agents and NLP pipelines with LangChain — designed for accuracy, traceability and cost control (model routing, semantic caching, structured outputs, human-in-the-loop).
- **AWS (deep):** Serverless and event-driven architecture — Lambda, API Gateway, EventBridge, SQS/SNS, DynamoDB, S3, Cognito, Bedrock, CDK — built for scale and sub-200ms latency.
- **Backend:** REST/GraphQL APIs and microservices in Python (FastAPI) and Node.js (NestJS/Express), with relational/SQL and NoSQL databases.
- **Fullstack:** React/TypeScript front ends, including micro-frontend architectures, on top of the services I build.
- **Also:** Python / TypeScript / JavaScript as primary languages; hands-on Microsoft Azure; working knowledge of Java for service integration.

TECHNICAL SKILLS

AI / LLM: Generative AI, RAG (semantic + hybrid search), LLM agents & orchestration, NLP document pipelines, Prompt Engineering, Model routing & cost optimization, Evaluation & guardrails, LangChain, LangGraph, Vector DBs (Pinecone, Qdrant), OpenAI, Claude, Gemini, AWS Bedrock

AWS (Cloud): Lambda, API Gateway, EventBridge, SQS, SNS, DynamoDB, S3, Cognito, Bedrock, CloudFormation, CDK — serverless & event-driven architecture at scale

Backend: Python (FastAPI), Node.js (NestJS, Express), Relational/SQL design (PostgreSQL, MySQL), MongoDB, Redis, REST APIs, GraphQL, Microservices, Java (integration)

Frontend: React, TypeScript, JavaScript (ES6+), Micro-Frontends, HTML5, CSS3, WebSockets

DevOps & Other Cloud: Git, GitHub Actions, CI/CD, Docker, Serverless Framework, Kafka; Microsoft Azure (Functions, Blob Storage, AI Search — hands-on)

Practices: Clean / Onion Architecture, SOLID, DDD, Event-Driven Design, TDD, Agile/Scrum, Code Review

PROFESSIONAL EXPERIENCE

Senior AI Engineer — Backend & Fullstack — Coderio

Feb 2024 – Present · Latin America (Remote)

Project: Coca-Cola Andina — AI-Driven Automation Platform

- Delivered full-stack features end to end: React/TypeScript dashboards backed by Python and Node.js services on AWS serverless infrastructure.
- Built an LLM observability dashboard (React) to monitor cost, latency and quality across 50+ daily AI workflows.
- Architected multi-agent systems with LangChain and AWS Bedrock, integrating GPT-4, Claude and Gemini.
- Built RAG pipelines with vector databases (Pinecone) and semantic search for context-aware answers over 100K+ ingested documents, including NLP cleaning and structuring.
- Designed a prompt-engineering framework that reduced token usage by 45% while improving response quality.
- Deployed 12+ serverless microservices (Lambda, API Gateway, SQS, SNS, EventBridge) with sub-200ms response times.
- Worked hands-on with Microsoft Azure services alongside AWS within the project's cloud environment.

Tech: React, TypeScript, Python, Node.js, NestJS, LangChain, AWS Bedrock, Microsoft Azure, AWS Lambda, DynamoDB, EventBridge

Senior AI Engineer — Backend — Distillery

Mar 2023 – Feb 2024 · Argentina (Remote)

Project: Naranja X — Financial Services Platform

- Built customer-facing conversational AI interfaces (React) plus the supporting Python/Node.js APIs, reaching a 70% inquiry deflection rate.
- Implemented a RAG system for financial document Q&A processing 50K+ queries monthly.
- Created AI-driven data-extraction pipelines from PDFs and unstructured documents using GPT-4.
- Deployed event-driven backend services on AWS with retry logic and exponential backoff for resilient LLM calls.

Tech: React, TypeScript, Node.js, Python, LangChain, OpenAI, AWS Lambda, S3, SQS, SNS, CDK

Lead Software Engineer — Fullstack & AI — NTT DATA Europe & LATAM

Mar 2022 – Aug 2023 · Lima, Perú

Project: AUNA Salud y Seguros — Healthcare Platform

- Led a team of 5 developers across front-end and back-end, owning delivery, code reviews and CI/CD pipelines.
- Pioneered LLM integration for automated patient communication and appointment scheduling, cutting administrative overhead by 50%.
- Built NLP pipelines to extract structured data from medical records and insurance documents.
- Architected onion architecture for clean separation of business logic and integrations.
- Hands-on with Microsoft Azure services as part of the platform's cloud infrastructure.

Tech: TypeScript, Node.js, Python, React, LangChain, OpenAI, Microsoft Azure, AWS Lambda, DynamoDB, Cognito, CDK

Software Engineer — Backend & Team Lead — COMPLEXLESS

Mar 2021 – Feb 2022 · Lima, Perú

Project: Unified Social Networks API (YouTube, Facebook, Twitter, Instagram)

- Built a unified API integrating multiple social platforms on an AWS serverless architecture.
- Led and mentored a team of junior developers — task planning, code reviews, and internal training on Node.js, micro-frontends and AWS.
- Implemented consumption and cost metrics across AWS services.

Tech: Node.js, JavaScript, AWS Lambda, DynamoDB, SNS, Cognito, CloudWatch, Serverless

Software Engineer — Backend — Convoca / Tismart

2021 – 2022 · Lima, Perú

- Convoca (DeepData): built Node.js/Express services with Sequelize and PostgreSQL; requirements estimation and feature delivery.
- Tismart (Bellcorp): built AWS Lambda functions (TypeScript/JavaScript) for email and event handling across API Gateway, S3, SQS, SNS, Cognito.

Tech: Node.js, Express, TypeScript, JavaScript, PostgreSQL, MongoDB, Kafka, Docker, AWS

Software Engineer — Backend — FuturaSoft · IDW (earlier roles)

2020 – 2021 · Tacna, Perú

- Gathered requirements with stakeholders, designed data/flow diagrams, and built Node.js/Express services.
- Automated business processes and delivered backend features with MySQL/MongoDB and Sequelize.

Tech: Node.js, Express, JavaScript, MySQL, MongoDB, Sequelize

IT Assistant — Industrias San Miguel

Nov 2018 – Jun 2019 · Tacna, Perú

- Ensured functionality of commercial systems across 9 branches in southern Peru; IT support and help desk.

Administrative & IT Assistant — Municipalidad Provincial Jorge Basadre

Jan 2018 – Jul 2018 · Ite, Perú

- Supported administrative processes, project execution and IT systems.

Software Developer — Universidad Privada de Tacna

Feb 2014 – Jul 2017 · Tacna, Perú

- Developed administrative and pedagogical systems for the university.

Developer — JyC Asesores y Consultores SAC

Sep 2012 – Feb 2013 · Perú

- Software development for administrative and business systems.

EDUCATION

Bachelor in Systems Engineering — Information Security · Universidad Privada de Tacna · 2007–2013

CERTIFICATIONS & LANGUAGES

- **Certifications:** SQL & Node.js (Talently), Web Front-End, Python (Data Structures & Control Structures).
- **Languages:** English — Professional Working Proficiency · Spanish — Native.