

JOAN PELLEGRINO

AI / GenAI Engineer | Multi-Agent Systems | RAG | Distributed Systems

Monteria, Colombia (Remote) | +57 310-409-6239 | joanpellegrino04@gmail.com | joanpellegrino.com |
linkedin.com/in/joanpellegrino

SUMMARY

AI/GenAI Engineer and Senior Integrations Specialist with nearly 6 years of experience building production-grade AI applications, data pipelines, and distributed systems. Specialized in stateful multi-agent orchestration (Router + Specialists), RAG pipelines, secure Text-to-SQL agents, and LLM developer tooling using Google ADK, LangGraph, MCP, Python, and cloud services (AWS & GCP). Combines applied AI application design with strong foundations in event-driven architectures, high-volume financial data pipelines, and 99.9% SLA enterprise infrastructure.

AI / GENAI TECHNICAL SKILLS

Agentic AI: Google ADK, LangGraph, LangChain, Multi-Agent Systems (Router + Specialists), Agent Orchestration, Tool-Using Agents, MCP (Model Context Protocol), Claude Agentic Skills

LLM & Real-Time Applications: RAG, Text-to-SQL, Gemini API (incl. Gemini Live API), Claude API, OpenAI API, Vector Databases (ChromaDB), Semantic Extraction, Real-Time WebSockets Audio Streaming

AI Safety / Controls: Strict Schema Validation, Read-Only SQL Query Controls, Session State & Memory Persistence (TTL), Contextual Result Explanations

Engineering: Python, FastAPI, PySpark, Pandas, SQL, TypeScript, REST APIs, Webhooks, Docker, Kubernetes (AWS EKS), CI/CD, Git

Cloud / Data: AWS (EKS, Lambda, S3, API Gateway, DynamoDB, SQS, Batch, CloudWatch), GCP (BigQuery, GCS, Vertex AI), Apache Kafka, Snowflake, PostgreSQL, ETL/ELT

PROFESSIONAL EXPERIENCE

SIMETRIK INC. - Senior AI & Integrations Engineer

Dec 2020 - Sep 2026 | Colombia (Remote)

- **Architected and deployed a stateful multi-agent system on AWS EKS** using **Google ADK and LangGraph** (Router + Specialists pattern) to autonomously resolve operational queries, troubleshoot CloudWatch logs, and orchestrate job re-executions in AWS Batch with persistent memory in **DynamoDB**.
- **Built a conversational Text-to-SQL agent** connected to **PostgreSQL via MCP (Model Context Protocol)**, enforcing strict schema validation, read-only security controls, and contextual result explanations for business teams.
- **Designed and delivered a semantic extraction and RAG pipeline** over technical documentation and unstructured contracts, enabling non-technical teams to query specialized knowledge bases via natural language and reducing operational reliance on specialists.
- **Created the Smart-Parser Claude Plugin** with specialized agentic skills in **Python** to automatically generate, validate, and audit complex data parsers and ETL scripts, drastically accelerating data ingestion and Git integration cycles.
- **Architected and implemented enterprise ETL/ELT pipelines** across REST APIs, SFTP, AWS S3, GCS, BigQuery, PostgreSQL, and **Snowflake**, processing **millions of financial transactions daily** with **99.9% availability (SLA)**.
- **Implemented fault-tolerant, event-driven architectures** using **Apache Kafka, webhooks, AWS Lambda, API Gateway, and Amazon SQS (DLQ)**, paired with high-performance validation engines in **Python and PySpark** for financial datasets (MT940, structured PDFs, CSV, and Parquet).

AI / GENAI ENGINEERING PROJECTS

Enterprise Multi-Agent Intelligence System (AWS EKS Orchestrator)

Python | Google ADK | LangGraph | MCP | RAG | Text-to-SQL | AWS EKS | DynamoDB

- Central intent router (gemini-2.0-flash) classifying incoming events (Slack, Zendesk, CLI) and delegating execution across isolated specialist agents (SQL Query Agent, Documentation RAG Agent, Diagnostic Agent, and Re-execution Agent).

Smart-Parser Claude Plugin & Real-Time Voice AI Agent

Python | Claude API | Gemini Live API | WebSockets | Web Audio API | TypeScript | Astro

- **Smart-Parser Plugin:** Declarative agentic engine for Claude Code CLI that synthesizes compliant ETL connectors, performs automated security/structural reviews, and automates Git Pull Requests.
- **Voice AI Agent (Live at joanpellegrino.com):** Real-time bidirectional voice agent powered by Google Gemini Live API over WebSockets and Web Audio API with a reactive state machine.

EDUCATION & CREDENTIALS

Undergraduate Degree in Systems Engineering / Computer Science - Universidad de Cordoba | 2013 - 2019 | Colombia

Thesis: Laureate IoT control and monitoring system | **Certification:** Python 3 Programming - University of Michigan | **Recognition:** IBM Behind the Code Marathon - AI & Data

Languages: Spanish (Native) | English (Technical Reading & Documentation)