

LUIS DAVID RUIZ SANTIAGO

AI Data Scientist

ldruizsan@gmail.com - (220)-238-4738

[LinkedIn](#) - [GitHub](#) - [Tableau Public](#) - [Portfolio](#) - [Hugging Face](#)

SUMMARY

AI Engineer with hands-on experience building end-to-end LLM applications, including a production RAG chatbot (ingestion, chunking, embedding, vector DB retrieval, and response generation) deployed on HuggingFace Spaces with guardrails and rate-limiting to manage cost and abuse. Also developed a no-code agentic workflow integrating API data retrieval, Python-based processing, and automated email reporting. Skilled in designing reliable, cost-conscious AI systems from prototype to deployment.

TECHNICAL SKILLS

Cloud & Automation: AWS, Azure, Python, PowerShell, Terraform (IaC), Docker

AI & Workflow Automation: n8n, Lindy, Agentic AI, RAG, prompt engineering, Claude, OpenAI, Gemini

Collaboration Tools: Jira, Confluence, Agile Methodologies

Data Analysis: Excel, Python (Pandas, Numpy, Matplotlib, Plotly,), Statistics

SELECT PROJECTS

Digital Twin: Conversational AI Persona Chatbot - [Digital Twin](#)

- Architected a "Digital Twin" chatbot using a RAG (Retrieval Augmented Generation) pipeline to emulate my persona based off my general profile, experience, and education with high accuracy
- Implemented a persistent vector database using ChromaDB and use OpenAI embeddings model to enable semantic search and context injection to reduce hallucinations
- Integrated tool function calling to extend capabilities, allowing the bot to perform external actions like sending messages to myself on behalf of the user, while also setting guardrails in place to prevent abuse and control costs
- Deployed a responsive web interface on Hugging Face Spaces using Gradio, managing secure environment variables, monitoring RAG retrieval latency, LLM response latency, and overall performance of the application, and API rate limiting

EPA Air Quality Data Collection AI Agent Workflow - [Lindy Agent Workflow](#)

- Built an AI agent workflow with Lindy to automate the collection of air quality data from the EPA website using their API
- Agent can be triggered adhoc or follow a set schedule with a human-in-the-loop guardrail to confirm and authorize the data collection process
- When the task is done, the agent exports the data onto a Google Sheets spreadsheet and sends an email with a summary of results and a link to the worksheet.

Pet Medication Scheduler- [website link](#)

- The pet medication scheduler is an application built with AI assistance that lets pet owners set, manage, and visualize the dosage calendar for medications
- Features include input of pet information, setting medication details like dose and duration, clearing or updating medication data, downloading a PDF version, and providing visual feedback to let the user know of any errors
- Stack: Python + Flask, HTML, CSS, Javascript, Heroku

EXPERIENCE

AI Fellow - SuperDataScience

Jun 2026

- AI Engineering cohort program designed to build a community of like minded individuals becoming AI engineers
- Built a conversational LLM-powered chat bot to mimic my persona, with a RAG pipeline, ChromaDB as vector database, and deployed to HuggingFace Spaces

Technology Architect Analyst - Accenture, Columbus, OH

Jun 2022 - Dec 2024

- Staffed across client engagements in digital transformation, cloud migration, and systems modernization for clients in the finance, fast food, and retail sectors
- Spearheaded documentation and testing frameworks for a multiplatform SaaS application, including authoring dozens of user stories for QA teams, creating a Product Playbook detailing feature entry points, cross-platform workflows, and ideal user paths to drive testing efficiency, and developing a Testing Matrix to standardize post-release validation to reduce duplication risk. Progress was tracked using Jira
- Led data validation and cleanup for legacy system modernization, ensuring accuracy for cloud migration, and identification of data gaps by comparing transaction logs to application outputs
- Designed a data workflow diagram and revamped training documentation to accelerate the onboarding of client teams
- Managed client-facing slides for sprint progress, risks, and blockers to improve transparency and maintain communication
- Collaborated with infrastructure teams to analyze system usage patterns and optimize SLAs, reducing downtime
- Ensured client data and company data security by staying up to date with security policies and practices

Fulfillment Center Associate I - Amazon CMH4, West Jefferson, OH

Nov 2025 - Jun 2026 | Sep 2021 - Apr 2022

- Ensure fast and accurate transfer of 2000-2500 items per day from inventory to packing with a low DPMO count
- Stow items to add them to inventory, perform quality tests to ensure accurate item count, and pack the items securely for shipment
- Follow 5S guidelines to maintain a clean and safe workstation

CERTIFICATIONS

- **Certified Agile Professional** (Six Sigma Global Institute), 2026
- **Data Modeling in Snowflake & dbt** (Datacamp), 2025
- **AI Agents and Agentic AI in Python** (Vanderbilt University), 2025
- **IBM Data Analyst Professional Certificate**, 2020
- **Data Analytics Certificate** (Thinkful), 2021
- **Certified Lean Six Sigma Yellow Belt Professional** (Six Sigma Global Institute), 2021
- **AWS Certified Cloud Practitioner CLF-01**, 2024
- **Azure Fundamentals AZ-900 & AI900**, 2024
- **Databricks Lakehouse Fundamentals**, 2024

EDUCATION

University of Akron

Sep 2016 - Oct 2018

Doctorate Coursework in Polymer Engineering (No degree completion)

University of Rochester - River Campus

Sep 2013 - May 2015

Alternative Energy - Chemical Engineering, M.Sc.

University of Puerto Rico - Mayaguez

Aug 2007 - Jun 2013

Chemical Engineering, B.Sc.