

# SAM KULKA

Austin, TX 78704 | (719) 487-5416 | [samkulka98@gmail.com](mailto:samkulka98@gmail.com) | [LinkedIn](#)

## AI-DATA ENGINEER | CLOUD & AI-DRIVEN AUTOMATION SPECIALIST

### SUMMARY AND PROFILE

---

AI-focused Data Engineer with expertise in building intelligent data systems that fuel machine learning, agentic workflows, and real-time analytics. Experienced in designing end-to-end pipelines that automate ingestion, cleansing, and feature engineering for AI deployment. Combining deep ETL automation skills with model-ready data design, enabling teams to deliver faster insights, scalable AI products, and continuous improvement through feedback-driven pipelines.

### CORE COMPETENCIES

---

• SQL • Python • R • Java • C++ • GCP • AWS • Agentic Engineering • Kubernetes  
• Data Orchestration • AI/ML Agile/Scrum • Informatica • Databricks • dbt • BigQuery • DuckDB  
• PostgreSQL • ETL Automation • Data Modeling • Data Warehousing • Containerization • IaC  
• Docker • Tableau • BI Tools

### EXPERIENCE

---

#### **Charles Schwab** *Data Engineer*

*Jan 2024 - May 2025*

- Led end-to-end design and delivery of a production data product, translating ambiguous regulatory and business requirements into a scalable system used by non-technical stakeholders to make high-stakes decisions.
- Led an acquisition and data warehouse migration initiative that saved Charles Schwab \$180M upon completion of the project.
- Engineered resilient data architectures supporting real-time analytics and predictive insights used by 200+ analysts across finance and operations.
- Implemented end-to-end data quality frameworks and validation logic, elevating accuracy and regulatory compliance across finance datasets while validating 5M records monthly with accuracy.

#### **Charles Schwab** *Data Engineer Intern*

*Jun 2022 - Aug 2023*

- Assisted in developing and tuning ETL pipelines using Informatica PowerCenter/IICS to streamline cross-source data integration.
- Enhanced ETL efficiency with SQL and Python scripts enabling faster query response times, cutting average query time by 40%.
- Automated 15+ workflows across 3 business units using data orchestration tools and GCP.
- Collaborated on sprint planning and backlog grooming to refine technical requirements and maintain agile velocity.

### PROJECTS

---

#### **Lex — AI Knowledge & Workflow Platform (Personal Project) - 2025**

- Built a full-stack AI platform that converts unstructured documents and APIs into structured data ontologies and queryable knowledge graphs using RAG, vector embeddings, and semantic indexing.
- Engineered a RAG pipeline combining OpenAI embeddings, a vector database, and PostgreSQL to support high-precision semantic search and multi-document reasoning.
- Implemented ontology-driven data models that normalize heterogeneous inputs into a unified, machine-readable schema.
- Developed a Next.js + FastAPI production system with authentication, job history, and audit trails for traceable, enterprise-grade AI workflows.

### EDUCATION

---

#### **Texas State University** - *Fall 2023*

*Bachelor of Business Administration - Finance*

#### **Austin Community College** - *Summer 2021*

*Associate of Science - Computer Science*

### CERTIFICATIONS

---

- Google Cloud Platform Associate Engineer
- Professional Scrum Developer™ I (PSD I)
- Dagster Essentials