

Akshit Rampershad

"Turning enterprise operations into scalable AI-native systems."

813-397-9209 | akshitrampershad@gmail.com | [LinkedIn](#) | [GitHub](#) | San Francisco, CA | Open To Relocate

EXPERIENCE

AI Engineer | Cambrian Lab, Pleasanton, CA

May 2026 – Present

- Automated a **2hr manual document review system into a 10sec AI workflow handling 29+ daily requests** by building an LLM pipeline with Claude Sonnet 4 for structured extraction, entity matching, and downstream impact analysis.
- Engineered an end-to-end **Product Change Notification (PCN) to Manufacturer Part Number (MPN) matching system** using a **4-leg hybrid retriever** (exact, fuzzy, prefix, vector) with an **LLM ranking pass** and a 3-band score+margin policy.
- Designed **confidence-gated decisioning over ranked candidates**, using score and margin thresholds to auto-confirm high-confidence matches and route ambiguous ones to human review, achieving **0 false auto-confirmations across 1k+ runs**.
- Achieved **0.36s match latency across 70k+ records**, embedding 4 manufacturing fields with OpenAI text-embedding-3-small (1536-dim) into a unified cosine **Qdrant** collection, using keyword payload indexes for filtered HNSW top-k search.
- Integrated with SAP to **automate downstream Bill of Materials (BOM) impact analysis**, building a recursive where-used engine with graph traversal and cycle detection that **eliminated manual tracing of assemblies and finished goods**.

Data Engineer | 22nd Century Technology Inc., McLean, VA

Sept 2025 – Mar 2026

- Led technical architecture and delivery of a **governed enterprise RAG platform across 20K+ SharePoint documents**, translating manual federal/state workflows into a scalable AI-powered knowledge and proposal drafting system.
- Engineered a hybrid ingestion pipeline using Python, PySpark, and Graph API, indexing **distributed SharePoint content into a 1,024-dim Qdrant vector database** with **BGE-Large-EN-v1.5 embeddings** for section-aware document chunking.
- Designed LlamaIndex-based retrieval pipeline **narrowing 30 candidates to 8 through metadata filtering, vector search, and Cross-Encoder re-ranking**, with LangChain orchestration producing RFP-scoped drafts grounded in prior proposals.
- Built a held-out evaluation benchmark across 20 proposal queries/drafting prompts, achieved a **75%+ retrieval hit rate and 54% first-pass draft acceptance**, reducing draft assembly from **7 days to under 4 hours** across 20+ monthly proposals.
- Facilitated product delivery across 4 cross-functional teams for 8 sprints, developed AI adoption runbooks, and led enablement sessions for proposal writers, **achieved 88% sustained weekly adoption rates and 2X faster review cycles**.

Data Engineer | Surge Infotech, Broadlands, VA

Sept 2024 – Sept 2025

- Architected **high throughput geospatial pipelines feeding an ML site suitability model** on Databricks AutoML+MLflow, scoring parcel buildability against zoning rules and geotechnical restraints at **72% accuracy vs site feasibility assessments**.
- Engineered **FastAPI data extraction services ingesting 50M+ parcel/zoning records into AWS S3**, leveraging Databricks Auto Loader for incremental file discovery & schema inference and DLT for governed ingestion & automated quality gates.
- Designed and orchestrated a multi-stage Medallion data architecture (Bronze/Silver/Gold)** on Databricks LakeFlow, establishing automated schema evolution and real-time quality validation gates across **spatial and zoning** records.
- Developed an **OpenAI GPT-4 reporting engine that synthesized pipeline, model, market metrics into executive-ready reports**, accelerating strategic planning by 34% and informing AI roadmap priorities and quarterly financial decisions.

Business Data Analyst | UKG, India

Oct 2020 – Aug 2022

- Led business and technical requirements analysis for UKG TALK post-acquisition, **identified key product gaps and shaped team's strategic product roadmap and deliverables**, that supported 10,000+ active enterprise users worldwide.
- Conducted SQL funnel analysis and Python cohort segmentation to identify onboarding friction and **executed process optimization techniques that improved premium feature monthly conversion rate by 29% and trial to paid rate by 25%**.
- Developed **Jira defect tracking dashboards** for production and staging issues, enabling 4 cross functional Agile teams to resolve dependencies and critical bugs, driving a **43% improvement in phased sign-offs** and a **94% sprint delivery rate**.
- Delivered interactive **Tableau and Power BI dashboards** using advanced parameters and calculated fields, translating **campaign ROI** and segment conversion metrics into product and business decisions for **executive quarterly reviews**.
- Managed **cross-functional UAT loops, post-production checks, and product marketing launch documentation** to control critical integration friction for Fortune 500 clients, **contributing to a 2X increase in platform adoption within 12 months**.

EDUCATION

- **MS in Business Analytics and Information Systems**, University of South Florida, Tampa, Florida **Aug 2022 – May 2024**
- **BS in Computer Science and Engineering**, Gandhi Institute of Technology and Management, India **Aug 2016 – July 2020**

TECHNICAL SKILLS

- **Agentic AI & LLM Orchestration:** Anthropic Claude API, OpenAI GPT-4, Gemini-3.7-flash, LangChain, LlamaIndex, MCP, Prompt Engineering, Data Extraction, Tool Orchestration, Agent Orchestration, Claude Agent SDK, Multi-System Agents.
- **RAG & Retrieval:** Qdrant, pgvector, Hybrid Retrieval (BM25/dense/RRF), Vector Search (HNSW/ANN), 2Stage Retrieval, Embedding, Entity Resolution, Cross-Encoder Rank, Metadata Filtering, Payload Indexing, Chunking.
- **AI Reliability & Evaluation:** Confidence Calibration, Score/Margin Thresholding, Validation Harnesses, False-Positive Control, Human-in-the-Loop (HITL), Agent Tracing/Observability, LangSmith, Token/Cost Tracking, Latency Benchmarking.
- **Engineering & Cloud:** Python, PySpark, Pydantic, SQL, Java, FastAPI, Databricks (Auto Loader, DLT, Lakeflow, Medallion Architecture), Snowflake (Snowpipe, Streams & Tasks, Snowsight), MySQL, AWS S3/EC2, MLflow, AutoML, Git/Bitbucket.
- **Technical Delivery & Product Management:** Technical Feasibility Analysis, Product Gap Analysis, Enterprise Application Integration, POCs, Solution Delivery, Project Management, Stakeholder Management, Technical Documentation.

ACCOMPLISHMENT

- **Certifications:** Model Context Protocol ([Anthropic](#)), Claude with Amazon Bedrock ([Anthropic](#)), Claude with Google Cloud's Vertex AI ([Anthropic](#)), Building with Claude API ([Anthropic](#)), Introduction to Agent skills ([Anthropic](#)), Introduction to Subagents ([Anthropic](#)), Snowflake Data Engineering ([Snowflake](#)), McKinsey Forward Learning ([Mckinsey.org](#)).
- **Recognition:** Exceptional Performer Award (UKG -2022), Technical Excellence Awards (UKG -2021).
- **Leadership:** Graduate Assistant (USF-2024), Data Science Club (USF-2023), PMI Volunteer (Tampa-2022).