

Shloak Aggarwal

San Jose, CA | (408) 210-6400 | aggarwalshloak@gmail.com | linkedin.com/in/shloakaggarwal | shloakaggarwal.com

SUMMARY

AI engineer completing an M.S. in Artificial Intelligence at SJSU, currently building a RAG-based literature classification system as a Graduate Research Assistant. Four years of production engineering experience underpin the work, including a multi-agent LangGraph pipeline with measured hallucination reduction and an LLM-backed enterprise help desk with role-based access.

EXPERIENCE

San Jose State University | Graduate Research Assistant

June 2026 – Present

- Implement a candidate-paper relevance classification system using **ChromaDB for hybrid retrieval and an LLM for judging**, validating the model with a stratified, hand-labeled ground-truth labeling workflow, resulting in more accurate paper relevance selection for research reviews
- Compiled a **70,000-paper transportation literature corpus** with Scopus API enrichment pipeline, then organized the set under a 370-term taxonomy across 4 research domains for downstream retrieval for research team

King Children (3D eyewear e-commerce startup) | Senior Software Engineer

July 2024 – Jan 2025

- Founding engineer on a new B2B partnership portal (Next.js 14, TypeScript, GraphQL that pivoted King Children from direct-to-consumer to B2B, building it **from PRD to production MVP in 10 weeks** across roughly 8 portal routes and 40+ components, leading to **2 new biz partnerships** after launching at Vision Expo East 2024
- Architected a Notion-based CRM integration using a webhook that auto-populates design-task deadlines across 3 Notion databases, plus a process-tracking dashboard **optimized from an N+1 query bottleneck to a single batched, lookup-mapped query**
- Maintained the team's Alembic migration history on a live production PostgreSQL schema with about 20 migrations over 3.5 years, each staging-verified and peer-reviewed before release **with no production incidents**
- Authored **5+ product requirement documents** after engaging with cross-functional stakeholders and B2B partners and customers to scope new features

King Children (3D eyewear e-commerce startup) | Software Engineer

July 2021 – July 2024

- Built and owned an internal order-management console (React, Redux, Apollo/GraphQL) used daily by the fulfillment team: a paginated, multi-facet-filterable orders table with selection-driven bulk status, material, and date editing, **refactored twice as primary author of its modules**
- Developed a data-automation tool that pulled design images from AWS S3 and measurements from PostgreSQL to auto-generate HTML comparison grids, **reducing a product designer's weekly effort from 2 hours to 10 minutes**
- Built the order "on-hold" workflow **end to end**: schema migrations, backend status logic, a React UI for staff, and a scheduled Slack summary job to keep the fulfillment team synced
- Shipped a discount-code management and batch-generation engine (React, Python/Tornado) with multi-currency support, **retiring one-at-a-time code creation** for the marketing and partnerships teams
- Built a reusable UIKit/SwiftUI bridging utility with navigation handoff, **letting a seven-year-old iOS app adopt SwiftUI incrementally**; used it to rebuild the Profile and Browse screens
- Delivered a Firebase Cloud Messaging push system end to end across the Tornado backend and Swift iOS client, **handling token lifecycle across login and logout**

IIIT Delhi | Research Intern

June 2020 – June 2021

- Designed and built a Neo4j knowledge graph from **10,000+ university course records**, including student profiles and time-versioned course history, to rank personalized learning pathways for an external ed-tech industry partner

Rocket.Chat | Software Engineering Intern

May 2019 – Aug 2019

- Shipped a GDPR-compliant contacts-sync module for the React Native client in a 2.4K-star open-source codebase, backed by unit and integration tests at **90%+ coverage** to preserve iOS/Android parity, and designed accompanying feature prototypes in Figma

PROJECTS

Multi-Agent Market Research Analyst (RAG Pipeline)

2026

- Designed a three-agent LangGraph pipeline over FAISS dense retrieval on a 195-document corpus; benchmarked four LLMs on 55 hand-labeled queries for **48x higher entity F1 (0.337 vs. 0.007)** and **hallucination down from 8.0% to 5.3%**. Served via FastAPI with per-agent Langfuse tracing

Medical Natural Language Inference with TinyClinicalBERT

2026

- **Fine-tuned TinyClinicalBERT** on the MedNLI dataset with PyTorch and Hugging Face, using custom training loops, evaluation metrics, and checkpointing to classify entailment between clinical statement pairs

Predictive Modeling of Flight Delays at SFO

2025

- Built an end-to-end classification pipeline over **135,000+ SFO departure records** to predict delays past 15 minutes; XGBoost led on F1 after feature engineering, importance analysis, and leakage controls

AI-Enhanced WiFi Sensing for Security

2026

- Building a **camera-free security-sensing system** with RuView DensePose and ST-GCN models over signal data from a network of ESP32-S3 sensor nodes

EDUCATION

San Jose State University | M.S. Artificial Intelligence

Dec 2026

- Relevant Coursework: Machine Learning, Deep Learning, Data Mining, AI & Data Engineering, Enterprise Software Platforms
- Won **Louie Barozzi Scholarship**, awarded annually to 1 international grad student for academic excellence and campus leadership
- **Treasurer** for NextGen | German Connect; **International Peer Mentor**

IIIT Delhi | B.Tech. Computer Science and Engineering

June 2021

SKILLS

Languages: Python, TypeScript, JavaScript, SQL, Java

AI & LLM: LangGraph, FAISS, Langfuse, ChromaDB, OpenAI API, Hugging Face, PyTorch, RAG, multi-agent orchestration, evaluation harnesses, prompt design

Backend & Data: FastAPI, Tornado, SQLAlchemy, Node.js, REST, GraphQL, PostgreSQL, DynamoDB, Auth0

Cloud & Practices: AWS (S3, ECS, CodePipeline), Vercel, Docker, GitHub Actions, CI/CD, code review, Agile

AI Developer Tools: Claude Code, Codex, Cursor, NotebookLM, Perplexity, CrewAI