Krish Shah

2003kshah@gmail.com
linkedin.com/in/krish-n-shah
609-556-8516

EDUCATION

Georgia Institute of Technology — Atlanta, GA
M.S. in Computer Science (OMSCS), Specialization in Computing Systems; GPA: 4.0/4.0 — *Jan 2026 – Present*

Boston University — Boston, MA
B.S. in Computer Engineering, Concentration in Machine Learning — *Sept 2022 – May 2025*

FIRST Robotics Presidential Scholar — top 0.5% of engineering admits; four-year half-tuition scholarship with funded research

EXPERIENCE

Clearwater Analytics (NYSE: CWAN) — New York, NY
Solutions Engineer — *June 2025 – Present*

- **Workflow Studio & Multi-Agent LLM Orchestration**: Architected orchestration platform managing 35+ LLM-powered agents (OpenAI, Azure OpenAI, AWS Bedrock) enabling 100+ researchers to automate commodity risk reports, deal analytics, and code generation. Reduced research cycle time by 16+ dev-hours per feature through AI-powered automation.
- **LLM Infrastructure & Multi-Provider Integration**: Designed infrastructure supporting embeddings, vector search, RAG for financial knowledge, and JSON-RPC MCP server with typed tool registry across 4 LLM providers. Implemented streaming support enabling seamless multi-model deployment and real-time agent responses.
- **Commodity Derivatives Greeks Analysis Platform**: Built full-stack platform (React + Python MVC) computing real-time Greeks (Delta, Gamma, Vega, Rho) across 1000+ commodity futures in < 2 minutes. Optimized query performance by 2–5×, enabling real-time risk dashboards for 10+ trading teams.

Kolochalama Laboratory — Boston, MA
Machine Learning Research Intern — *Sept 2024 – May 2025*

- **RAG Pipeline for Biomedical QA**: Built Retrieval-Augmented Generation system using Sentence-BERT embeddings and FAISS to ground LLM outputs in PubMed literature. Reduced hallucinations by 40% and achieved 98.5% citation accuracy for clinical AI responses.
- **Evaluation Framework & Benchmarking**: Designed evaluation framework measuring hallucination rate, citation accuracy, and semantic similarity (BERTScore). Established rigorous benchmarks for biomedical LLM reliability and factual grounding.

State Street — Boston, MA
Software Engineer Intern — *Sept 2024 – Dec 2024*

- **Snowflake CI/CD Automation Platform**: Built an automated Snowflake resource deployment platform using Liquibase, Harness CI/CD, and Python—supporting version-controlled migrations, RBAC provisioning, warehouse scaling, a reusable CLI tool, and full documentation adopted by external users.

Beacon Platform Inc (Pre-IPO, Series C) — New York, NY
Software Engineer Intern — *Sept 2023 – Sept 2024*

- **Equity Derivatives Accelerator**: Engineered Equity Derivatives Accelerator and Scenario Generation tool enabling rapid POC setup and analysis workflows. Delivered platform used by 70% of clients to reduce analysis setup time by 50%, directly contributing to accelerated client onboarding and product adoption across the platform.
- **AI Chatbot for RFI/RFP Automation**: Developed Llama-3-8B powered chatbot with document loading, text embedding, vector search, and RAG pipeline for semantic retrieval from internal documentation. Integrated response classification (good/bad/escalate) and deployed Slack bot reducing RFI turnaround by 40%.
- **Data Quality Framework**: Evaluated and implemented Great Expectations, created custom expectations library (Column Aggregate, Column Map, Regex-based, Query-based) with YAML configuration, automated testing, and Slack notifications. Produced internal documentation and knowledge transfer guides for data quality validation across commodity and financial datasets.

TECHNICAL SKILLS

- **Languages**: C++, Python, Java, MATLAB, TypeScript/JavaScript, SQL, VBA
- **Frameworks**: React, Svelte, Next.js, FastAPI, Node.js, Flask
- **Developer Tools**: Git, CI/CD, Docker, Kubernetes, Apache Airflow, Postman, AWS, GCP, PostgreSQL, Firebase
- **Libraries**: NumPy, Pandas, scikit-learn, PyTorch, TensorFlow, Seaborn, Matplotlib, Ray, FAISS, Hugging Face, LangChain

US Citizen