Krish Shah

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

Education

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

Relevant Coursework: Graduate Operating Systems; Computer Networks; Graduate Algorithms; GPU Hardware and Software; High-Performance Computing; Software Architecture and Design; Software Development Process; Machine Learning for Trading; Deep Learning; Data and Visual Analytics

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, funded research

Experience

Clearwater Analytics New York, NY
Solutions Engineer *June 2025 – Present*

- **Distributed Agent Orchestration & Execution Engine**: Built a distributed execution engine for 35+ LLM agents with streaming control flow, typed tool calls, and routing across Azure OpenAI and AWS Bedrock. Saved 16+ engineering hours per feature; runs commodity risk workflows for 100+ researchers.
- **LLM Infrastructure & Retrieval Systems (Async, Streaming)**: Built Tornado async APIs with WebSocket streaming, an OpenAI-compatible interface, and FAISS retrieval over financial knowledge bases. Also built a typed tool registry and JSON-RPC MCP server - schema validation, backpressure handling, retry logic - for production use.
- **Low-Latency Risk Analytics Engine (Full-Stack)**: Built a full-stack risk analytics engine (React + Python MVC) computing real-time Greeks (Delta, Gamma, Vega, Rho) across 1,000+ commodity futures in under 2 minutes. Query optimizations and caching cut latency 2–5×; used by 10+ trading teams.
- **Agentic Copilot for Excel**: Added an AI layer to Beacon by CWAN's Excel Add-In that reads and writes cells mid-conversation, pulls internal risk reports with one click, and lets users query live spreadsheet data in plain English via sidebar chat or native Excel formula calls.
- **Technical Demoware & Client Presentations**: One of three engineers on the NY team, working across London and Tokyo to build persona-driven demoware on Beacon's platform and present directly to 40+ institutional clients - risk managers, quants, and traders - throughout the sales cycle.

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

- **RAG Pipeline for Biomedical QA**: Built a RAG pipeline using Sentence-BERT embeddings and FAISS over PubMed literature. Cut hallucination rate by 40% vs. baseline; 98.5% citation accuracy on clinical QA.
- **Evaluation Framework & Benchmarking**: Built an evaluation framework measuring hallucination rate, citation accuracy, and BERTScore - made results reproducible and comparable across biomedical LLM runs.

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

- **Equity Derivatives Accelerator**: Built an equity derivatives accelerator and scenario generation system; 70% client adoption, 50% faster analysis setup. Covered data ingestion, scenario generation, and analytics output end-to-end.
- **Data Quality Framework**: Built a configurable data quality framework on Great Expectations with custom expectation libraries (aggregate, map, regex, query-based). YAML-driven config, automated testing, and Slack alerts across commodity and financial datasets.

Projects

- **High-Performance FFT for Financial Signal Analysis**: Wrote a radix-2 iterative FFT in C with OpenMP; 8.3× speedup on 16M-point inputs via cache-aware memory layouts, loop fusion, and in-place computation ($< 10^{-11}$ numerical error). Applied to AAPL and SPY log-returns for frequency-domain volatility and microstructure analysis.

Technical Skills

- **Languages**: Python, C++, C, SQL, MATLAB, Java, TypeScript/JavaScript
- **Systems & Infra**: Linux, perf, flamegraphs, CI/CD, Docker, Kubernetes, Airflow, AWS, GCP, PostgreSQL, Snowflake
- **Frameworks**: React, Next.js, FastAPI, Flask, CUDA, OpenMP
- **ML & Data**: PyTorch, TensorFlow, NumPy, Pandas, scikit-learn, FAISS, Hugging Face, Ray

US Citizen