

Rezwan Sheikh



(917) 463-8110

rezwanswe23@gmail.com

LinkedIn

GitHub

Education

St. John's University

Bachelor of Science in Computer Science (GPA: 3.6 / 4.0)

Expected Dec 2026

Queens, New York

- Relevant Coursework: Object-Oriented Programming, Data Structures, Database Management Systems

Experience

Momen Technologies

May 2025 – Present

Orlando, FL (Remote)

Lead Software Engineer

- Developed a knowledge assistant platform powered by an LLM to provide context-aware responses to queries, reducing response latency by 40% and stress-tested to handle 10K+ queries/day
- Built a cross-platform mobile app with React Native, integrating Firebase Authentication for secure authentication and onboarding flow, reducing login time by 30% and supporting 500+ concurrent users across iOS and Android.
- Designed and implemented a real-time AI chat interface with smooth animations and persistent sessions, enabling sub-200ms response times

Momen Technologies

Feb 2025 – Present

Orlando, FL (Remote)

Founding Engineer

- Fine-tuned DeepSeek-Chat 7B using LoRA (Low-Rank Adaptation) on domain-specific Islamic texts, improving AI response relevance by 30% while reducing GPU memory usage and training time by 40%
- Implemented RAG with LangChain, leveraging MongoDB for structured knowledge storage and Pinecone for vector embeddings, improving factual accuracy in AI responses by 20% while minimizing hallucinations
- Deployed a scalable NestJS backend on AWS EC2, reducing API response times by 35% through optimized load balancing and efficient request handling

NEW EIC

Dec 2024 – Mar 2025

Boston, MA (Remote)

Software Engineering Intern

- Implemented a scalable interview management platform with cloud-based file storage, version control, role-based permissions, and calendar integration using MongoDB, AWS S3, Google Calendar API, enabling 3x faster document retrieval, reducing backend storage size by 70%, and supporting seamless scheduling for 500+ student interviews
- Architected a strategy management system with RBAC, real-time collaboration, approval workflows, and audit trails, reducing strategy approval turnaround by 40% and enabling scalable oversight across 1,000+ submissions

Institute for Machine Learning

Sep 2024 – Dec 2024

Queens, NY

Software Engineering Intern

- Developed cloud-based infrastructure using AWS EC2 for scalable processing, AWS RDS for high-performance storage, and AWS Lambda for serverless compute, optimizing system performance
- Improved system scalability by 40% and reduced data processing time by 25% by fine-tuning cloud infrastructure and optimizing services for peak performance
- Used Java to develop time series models for autoregressive analysis, integrate econometric/statistical modules, deploy news indicator analytics for real-time financial insights, and implement data fetching services

Projects

Horizon | React, Next.js, TailwindCSS, Plaid, Dwolla, Appwrite, Sentry

- Developed a financial web application using Plaid for bank account linking and Dwolla for secure fund transfers
- Implemented real-time error tracking with Sentry, reducing error resolution time by 50%
- Maintained a 98% transaction success rate, ensuring reliable and consistent fund transfers

Technical Skills

Languages and Software: Java, Python, JavaScript/TypeScript, HTML/CSS, Node, Express, Nest.js, React, React Native, Next.js, Firebase, MongoDB, PostgreSQL, Supabase, Appwrite, TailwindCSS

AI/ML: NLP, Prompt Engineering, Multimodal AI, LLM Fine-Tuning (LoRA/QLoRA), RAG, Agentic AI

Tools: IntelliJ, Eclipse, AWS, Docker, GitHub, Jira, VS Code, Postman, Sentry, OpenAI, Hugging Face, LangChain, Pinecone

Certifications: AWS Certified Solutions Architect - Associate | Amazon Web Services (AWS) | Feb 2025