





# Rezwan Sheikh

 (917) 463-8110    rezwanswe23@gmail.com    LinkedIn    GitHub

## Summary

Full-stack Software Engineer skilled in Next.js, React Native, Node, and TypeScript, with experience building secure and scalable web and mobile applications. Strong background in authentication, real-time features, API integrations, and system design. Experienced in developing AI-driven features using modern LLMs and RAG pipelines. Proven ability to ship production-quality projects, solve complex problems, and deliver end-to-end solutions.

## Education

<b>St. John's University</b> <i>Bachelor of Science in Computer Science (GPA: 3.6 / 4.0)</i>	Expected Dec 2026 Queens, New York
---	---------------------------------------

## Experience

<b>Momen Technologies</b> <i>Mobile Software Engineer</i>	May 2025 – Present Orlando, FL (Remote)
<ul style="list-style-type: none"><li>Built an LLM-powered knowledge assistant with context-aware responses, cutting latency by 40% and scaling to 10K+ queries/day</li><li>Developed a React Native mobile app with Firebase Authentication, reducing login time by 30% and supporting 500+ concurrent users</li><li>Implemented a real-time AI chat interface with smooth animations and persistent sessions, enabling sub-200ms response times</li></ul>	
<b>Momen Technologies</b> <i>Software Engineer</i>	Feb 2025 – May 2025 Orlando, FL (Remote)
<ul style="list-style-type: none"><li>Fine-tuned DeepSeek-Chat 7B with LoRA, improving response relevance by 30% and reducing memory and training time by 40%</li><li>Built a RAG pipeline with LangChain, MongoDB, and Pinecone, boosting factual accuracy by 20% and reducing hallucinations</li><li>Deployed a scalable NestJS backend on AWS EC2, reducing API response times by 35% through optimized load balancing and efficient request handling</li></ul>	
<b>NEW EIC</b> <i>Software Engineering Intern</i>	Dec 2024 – Mar 2025 Boston, MA (Remote)
<ul style="list-style-type: none"><li>Built a scalable interview management platform with cloud storage using MongoDB and AWS S3, achieving 3x faster retrieval and 70% lower storage usage</li><li>Architected a strategy management system with RBAC, real-time collaboration, and approval workflows, reducing approval time by 40%</li></ul>	
<b>Institute for Machine Learning</b> <i>Software Engineering Intern</i>	Sep 2024 – Dec 2024 Queens, NY
<ul style="list-style-type: none"><li>Developed AWS infrastructure (EC2, RDS, Lambda) to improve scalability and system performance</li><li>Increased scalability by 40% and reduced data processing time by 25% through cloud infrastructure optimizations</li><li>Built time series models in Java, integrated econometric modules, and deployed real-time financial analytics</li></ul>	

## Projects

<b>Horizon</b>   <i>React, Next.js, TailwindCSS, Plaid, Dwolla, Appwrite, Sentry</i>
<ul style="list-style-type: none"><li>Developed a financial web application using Plaid for bank account linking and Dwolla for secure fund transfers</li><li>Implemented real-time error tracking with Sentry, reducing error resolution time by 50%</li></ul>

## 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