

# Rahul Shah

San Francisco, CA | +1(516)979-5019 | [rshah107@asu.edu](mailto:rshah107@asu.edu) | [Portfolio](#) | [Github](#) | [LinkedIn](#)

## EXPERIENCE

|  |   |
|--|---|
| <b>Software Engineer</b><br><i>Solutions Unified LLC.</i>  | May 2024 – Present<br>Califon, NJ         |
| <ul style="list-style-type: none"><li>Designed a scalable <b>file upload pipeline</b> using pre-signed <b>S3</b> URLs with chunked uploads, S3 multipart support and <b>Lambda</b> triggers, reducing core service load and improving file processing latency by <b>35%</b>.</li><li>Automated multi-service (<b>Elastic Beanstalk, S3, DynamoDB</b>) deployment using <b>AWS CDK</b>, reducing customer environment setup time by <b>75%</b> through seamless integration with <b>CloudFormation</b>.</li><li>Developed an admin dashboard using <b>React, Tailwind</b>, and <b>Go</b> backend services to monitor real-time usage metrics and system logs, enhancing system observability and reducing debugging time by <b>60%</b>.</li><li>Built a <b>notification service</b> to process usage data from client environments via <b>SNS</b> and <b>SQS</b>, triggering Lambda-based metering and <b>Stripe</b> integration for billing, reducing errors by <b>25%</b>.</li></ul>  |   |
| <b>AI Full-Stack Developer</b><br><i>Enterprise Technology(Mastercard Foundation), Arizona State University.</i>   | Jan. 2024 – Jan. 2025<br>Tempe, AZ        |
| <ul style="list-style-type: none"><li>Architected three full-stack platforms serving over <b>10,000+</b> users to build, compare, and interact with various <b>LLMs</b> using <b>AWS Bedrock, OpenSearch</b>, and <b>Llamaindex</b> with audio and image support.</li><li>Engineered a <b>voice-enabled assistant</b> by integrating <b>Whisper</b> for transcription and <b>Amazon Polly</b> for speech synthesis, allowing seamless bi-directional audio interaction with LLMs for accessibility-focused use case.</li><li>Architected a modular backend system supporting plug-and-play integration of third-party <b>LLM</b> providers, enabling flexibility in model experimentation without downtime.</li><li>Created a <b>multi-agent</b> orchestration system using <b>LangGraph</b> and <b>FastAPI</b>, enabling chained LLM workflows for planning, memory, and tool use, reducing user interaction steps by <b>50%</b>.</li></ul>   |   |
| <b>Software Engineer II (Full-Stack)</b><br><i>Allegion (Schlage).</i>   | June 2020 – July 2023<br>Bengaluru, India |
| <ul style="list-style-type: none"><li>Drove the migration of an internal API from <b>REST</b> to <b>GraphQL</b>, optimizing query complexity and efficient data fetching, reducing the need for multiple API calls &amp; improving developer efficiency by <b>35%</b>.</li><li>Led the integration of an automated <b>CI/CD pipeline</b> using <b>Jenkins, Docker</b>, and <b>Kubernetes</b>, reducing deployment cycles by <b>40%</b> and implementing blue-green deployment for zero-downtime production updates.</li><li>Developed an onboarding workflow using <b>React</b> and <b>Redux</b>, integrating REST APIs and custom hooks, improving task completion rates and reducing onboarding time by <b>60%</b>.</li><li>Built a middleware-based request tracing system in <b>Node.js</b> using OpenTelemetry, enabling distributed tracing across microservices and reducing debugging time by <b>40%</b>.</li><li>Designed and implemented an <b>end-to-end testing</b> framework using <b>Selenium WebDriver</b> and <b>Mocha</b>, enabling automated browser testing reducing manual QA efforts by <b>70%</b>.</li></ul> |   |

## PROJECTS

|   |                      |
|---|----------------------|
| <b>Conversense</b> ( <a href="#">link</a> )   <i>Go, React, PostgreSQL, LiveKit, Google Gemini, AWS S3, Ingest</i>  | Oct 2025 – Nov 2025  |
| <ul style="list-style-type: none"><li>Built an AI-powered meeting platform with real-time video conferencing, automated transcription, and intelligent summarization using LiveKit and Google Gemini, featuring AI agents for in-meeting assistance and post-meeting chat interaction.</li></ul>                                      |                      |
| <b>Vistruct</b> ( <a href="#">link</a> ) ( <a href="#">Demo</a> )   <i>React, TypeScript, FastAPI, PostgreSQL, Docker, GenAI</i>  | Jul. 2025 – Aug 2025 |
| <ul style="list-style-type: none"><li>Developed a full-stack educational video platform that is currently used by 20+ professors from my undergrad university, leveraging AI for automated content generation, featuring real-time chat-based script development, AI voice-over synthesis, and Manim animation integration.</li></ul> |                      |
| <b>InfoFetchBot</b> ( <a href="#">link</a> )   <i>Python, Discord API, Open AI, AutoGen, NotionAPI</i>  | Sep. 2024 – Oct 2024 |
| <ul style="list-style-type: none"><li>Created a Discord bot that processes incoming messages, triggers web searches, and scrapes website content using AI agents, delivering concise summaries directly into Notion for data organization and retrieval.</li></ul>  |                      |

## TECHNICAL SKILLS

**Languages & Frameworks:** Go, Java, TypeScript/JavaScript, Python, FastAPI, React, Node.js, Spring Boot  
**Cloud & Tools:** AWS (EC2, API Gateway, S3, ECS, Bedrock), Kubernetes, Docker, Git, Jenkins, Selenium  
**Data & Messaging:** DynamoDB, PostgreSQL, MySQL, MongoDB, Redis, Kafka, RabbitMQ, SQS, SNS

## EDUCATION

|   |                                   |
|---|-----------------------------------|
| <b>Arizona State University</b><br><i>Master of Science in Computer Science</i> <b>GPA: 4.0/4.0</b> | Tempe, AZ<br>Aug. 2023 – May 2025 |
|---|-----------------------------------|