Smit Lila

(562)912-0591 | smitlila07@gmail.com | linkedin.com/in/smitlila | github.com/Smitlila

Education

California State University, Long Beach — **Dec 2025**

Bachelor of Science in Computer Science, GPA: 3.6
Certifications: AWS Cloud Practitioner, The MERN Full-Stack Guide (Udemy), Python Data Structures
Awards: CSULB President's Honor List

Work Experience

Software Engineer Intern — **May 2025 – Aug 2025**

Amazon Web Services — *Seattle, WA*

- Designed and deployed the **NXAutoAnalyzer Bedrock Agent**, an AI-driven automation framework that analyzes and posts **Root Cause Analyses (RCA)** for encryption-related incidents, reducing manual investigation time by **70%**.
- Integrated **Claude 3.7 Sonnet with AWS Bedrock** using the **Strands Agent SDK** to execute reasoning loops, leveraging real-time log queries, EC2 metadata retrieval, and runbook analysis under the **Model Context Protocol (MCP)** architecture.
- Implemented a **Lambda-based filtering service** to automatically trigger the AI agent for supported error types, increasing system scalability and reducing false triggers by **40%**.
- Built automated evaluation pipelines using **Multiple-Choice** and **Short-Answer** frameworks (**LLM-as-a-Judge**) to measure reasoning accuracy, achieving **97%+ correctness** on structured diagnostic tasks.
- Redesigned complex troubleshooting runbooks into **LLM-friendly, structured workflows** with defined queries and thresholds, improving **RCA consistency** and **diagnostic accuracy** by **35%**.
- Automated deployment and validation using **AWS Batch**, **S3 logging**, and **CI/CD pipelines**, ensuring repeatable, auditable, and fault-tolerant execution across hundreds of test tickets.
- Delivered a **production-ready AI system** integrating **data engineering**, **MLOps**, and **software automation**, paving the way for scalable AI diagnostics across multiple NX infrastructure services.

Software Engineer Fellow — **May 2024 – Aug 2024**

Headstarter AI — *Long Beach, CA*

- Built dynamic, responsive front-end applications using **React** and **TypeScript** enhancing performance and scalability.
- Developed a **customer analytics tool** with **Python** and **React**, reducing data processing time by **30%**.
- Collaborated with industry experts to create **AI solutions** and contributed to **open-source projects**.
- Gained hands-on experience with **AI tools and platforms**, including **Python**, machine learning frameworks, leading to the successful implementation of advanced AI features in applications.

Instructional Student Assistant — **Feb 2024 – Present**

California State University, Long Beach — *Long Beach, CA*

- Led **curriculum development**, boosting engagement by **15%** through interactive **Python workshops**.
- Provided one-on-one mentorship to students, improving their problem-solving and debugging skills.
- Trained new tutors on effective **teaching methods** and **participation strategies**, improving student outcomes.

Projects

Chat with PDF | *Generative AI RAG Application using AWS Bedrock* — **April 2025**

- Built a **two-app RAG system** (ingest + query) in **Python/LangChain** using **Claude** via **Bedrock** and **Titan Embeddings**.
- Enabled queries on **PDFs up to 500 pages** in **3 seconds** and sustained **50+ concurrent users** with **Docker on AWS EC2/S3** and **IAM**.
- Implemented **chunking**, **caching**, and **guardrails** to ensure relevant responses while maintaining low latency.

DevSecOps Project | *Netflix Clone Deployment on AWS* — **Feb 2025**

- Designed **Jenkins pipeline** with **Docker build/scan** (**SonarQube**, **Trivy**), blocking **15+ vulnerabilities** before deployment.
- Automated deployments with **Helm + ArgoCD** on **Kubernetes**, streamlining **GitOps workflows** and improving delivery speed.
- Integrated **Prometheus** and **Grafana**, enhancing system monitoring and reliability for **production use**.

Rate-My-professor AI | *JavaScript, React, Next.js, Tailwind, Gemini API* — **Oct 2024**

- Developed an **AI-powered professor-rating web app** with real-time assistance, improving user engagement by **20%**.
- Implemented a dynamic **UI** and integrated the **Gemini API** to handle **100+ requests/min**, reducing load times by **25%** and optimizing scalability.
- Enhanced authentication with **Pinecode** and refined **UX design**, improving data efficiency by **15%** and increasing user satisfaction by **30%**.

Technical Skills

Languages: Python, Java, JavaScript (ES6+), TypeScript, Java, C++, Objective-C, C#, Go, SQL

Web/App: React, Node.js, Express.js, Next.js, Tailwind CSS, Kotlin, SwiftUI, iOS, REST APIs, JWT, Socket.IO

AI/RAG: LangChain, AWS Bedrock (Claude, Titan Embeddings), OpenAI, TensorFlow, PyTorch, Scikit-learn

DevOps/Cloud: AWS (EC2, S3, EKS, Lambda), Linux, Azure, Docker, Kubernetes, Jenkins, GitHub Actions, Databricks

Tools/DB: MongoDB, PostgreSQL, Tableau, Power BI, Postman, Grafana, Git/GitHub