Subham Mitra

224.202.1423 | subham.mitra@berkeley.edu | linkedin.com/in/subham | github.com/subham | subham.wiki/

Education

University of California, Berkeley **Expected Spring 2026**

B.S. Electrical Engineering & Computer Science *GPA: 3.76/4.0*

Relevant Coursework: Data Structures & Algorithms, Efficient Programs, Machine Structures, Discrete Mathematics, Probability & Random Processes, Linear Algebra for ML, Intro to Artificial Intelligence, Digital Design & Integrated Circuits

Skills: Python, Java, C++, RISC-V, JavaScript, Swift, LlamaIndex, OpenAI GPT, TensorFlow, Pandas, RAG, Prompt Engineering, AWS, Terraform, GitHub Actions, CI/CD, React, Node.js, R, PostgreSQL, Vector Databases, Graph Databases

Extracurriculars: Blockchain @ Berkeley, Mobile Developers of Berkeley, 180 Degrees Consulting

Work Experience

Sensigo - Porsche Ventures **Jan 2025 – Present**

MLOps Engineer *San Fransisco, CA*

- Engineered AWS SageMaker infrastructure transition from serverless to provisioned endpoints, implementing load testing and batch processing capabilities to support expanding dealership network while reducing latency by 40%.
- Designed and implemented a K-Nearest Neighbors model for vehicle diagnostics using scikit and DVC pipeline with stages for data extraction, multihot encoding preprocessing, and custom distance metrics. Deployed production-ready inference system with normalization and filtering via AWS SageMaker endpoint, expanding repair location coverage by 150%.
- Developed Terraform-based multi-environment deployment system with feature flags and selective model updates, optimizing CI/CD workflow and reducing unnecessary model redeployments by 75%.

Oracle NetSuite **Nov 2024 – Present**

Machine Learning Engineer *Redwood City, CA*

- Developing intelligent workflow automation system that translates natural language into NetSuite SuiteFlow XML, utilizing fine-tuned LLaMA and Cohere models to understand business logic and generate compliant workflow diagrams.
- Architecting hybrid RAG pipeline that combines vector embeddings of XML templates with graph-based representation of workflow semantics, achieving 85% accuracy in converting business requirements into executable SuiteFlow graphs.
- Engineering custom semantic parsing system using domain-adapted language models to map business process descriptions to XML schema components, implementing context window optimization to handle verbose enterprise workflows.

SXS Unlimited Rentals *via Mobile Developers of Berkeley* **Sep 2024 – Present**

Technical Project Manager *Berkeley, CA*

- Leading team of 6 React Native developers in building $31k vehicle rental marketplace app, managing 11 milestone deliverables including vehicle management, booking systems, payment integration, and user role differentiation.
- Architected and deployed scalable serverless backend infrastructure using AWS CDK, Lambda functions, S3 storage, Aurora PostgreSQL, AppSync, and Cognito, reducing projected monthly costs by 80% through serverless compute.
- Established comprehensive end-to-end development workflow including code review processes, automated CI/CD pipelines, and structured biweekly sprints, consistently maintaining 100% on-time delivery across all project milestones.

International Computer Science Institute **Aug 2024 – Dec 2024**

Research Assistant *Berkeley, CA*

- Led statistical analysis for an 800-participant study on telehealth app privacy, conducting mixed-effects regression modeling (CLMM, GLMM) and implementing robust statistical tests including Wilcoxon and McNemar's tests.
- Executed comprehensive data transformation and validation procedures, including multicollinearity testing (VIF analysis), heteroskedasticity assessment (Breusch-Pagan), and distribution analysis (Kolmogorov-Smirnov), ensuring statistical rigor in analyzing non-normally distributed bounded proportion data across multiple dimensions.

UC Berkeley EECS Course Staff | CS 70 **Aug 2024 – Dec 2024**

Teaching Intern *Berkeley, CA*

- Lead weekly discussion sections for 700-person undergraduate course, implementing interactive teaching methodologies to convey concepts in probability theory, modular arithmetic, graph theory, and rigorous mathematical proof techniques.
- Provide individualized student support through one-on-one tutoring sessions, and answering 50+ questions in discussion forum, focusing on complex topics like counting principles, conditional probability, and discrete random variables.

Personal Projects

AI Lecture Assistant | *AWS Bedrock RAG, OpenAI GPT, Hume, React, Node.js* |

- Implemented AI system combining AWS Bedrock's RAG architecture for semantic search, OpenAI GPT for contextual quiz generation, and Hume's speech-to-text API for real-time transcription, for a cohesive lecture viewing platform.
- Engineered custom prompt engineering pipeline that processes YouTube video transcripts through vector embeddings, utilizing parallel API calls and context window optimization to maintain coherent responses across lecture segments.

Mentora | *React, Deepgram, Cartesia, OpenAI, Node.js* |

- Architected and developed an interactive voice AI tutoring platform using React and Node.js, integrating Deepgram's real-time speech analysis, Cartesia's voice cloning technology, and Tldraw's collaborative whiteboard interface.
- Engineered a multi-modal learning system that processes concurrent audio, visual, and text inputs through OpenAI API, implementing custom sentiment analysis to deliver adaptive, step-by-step educational guidance for K-8 students.