

# SHOHAIB MALLICK

Pune, Maharashtra | [shohaibsmallick@gmail.com](mailto:shohaibsmallick@gmail.com) | +91 7798872021 | [Portfolio Website](#) | [Linkedin](#)

## EDUCATION

**Boston University**  
**Master of Science, Computer Science (GPA 3.84 / 4)**

**Boston, Massachusetts, USA**  
Graduation Date: May 2024

**Savitribai Phule Pune University**  
**Bachelor of Engineering, Computer Engineering (GPA 8.48 / 10)**

**Maharashtra, India**

## WORK EXPERIENCE

**Aaum AI, Pune, Maharashtra (Startup, B2C)**

**Sept 2025 - now**

Full Stack Software Engineer

- Developed and integrated scalable APIs in **Node.js (TypeScript)**, implementing **Razorpay payment processor with webhooks** for a subscription-based model improving payment success rate and enabling seamless recurring billing.
- Designed and optimized backend systems using **PostgreSQL** and **Redis caching**, reducing query response time by **~35%** and enhancing overall API performance.
- Built responsive front-end interfaces with **React (TypeScript)** and **Tailwind CSS**, leveraging **Zustand** for efficient, predictable state management across complex UI workflows.
- Engineered an **agentic Retrieval-Augmented Generation (RAG)** service using **Agno AI, Python**, and **FastAPI**, integrating **Google Custom Search** for web search capability along with **LanceDB** for storing vector embeddings to enable **semantic search** and contextual retrieval.
- Integrated multiple **LLM providers (Perplexity, OpenAI, and Grok APIs)** for dynamic chat completions, enhancing response diversity, reliability, and contextual accuracy in AI-driven workflows.
- Deployed services using **Docker** on **AWS EC2**, ensuring scalable, fault-tolerant infrastructure

**FIND ME, USA (Startup, B2B2C) | [\[link\]](#)**

**Dec 2024 - July 2025**

Full Stack Developer

- Core contributor to MVP development**, collaborating closely with **founders** and **stakeholders** to define product requirements and deliver the first production-ready version under tight deadlines.
- Designed and deployed **RESTful APIs**, enabling **seamless real-time data integration** and supporting future **scalability**.
- Built **typo-tolerant fuzzy search**, increasing user discoverability and **reducing bounce rate** by **30%**, directly shaping the product's user experience.
- Optimized backend architecture, **reducing API latency** by **40%** and ensuring the **MVP** could support early **user growth** without performance bottlenecks.
- Integrated fault-tolerant data layers using **PostgreSQL, MongoDB, and Redis**, ensuring data durability under high concurrency.
- Developed centralized **observability** infrastructure (**OpenTelemetry, Promtail, Grafana Loki**), enabling rapid **debugging** and improving release confidence.
- Implemented secure, scalable **OAuth2/JWT authentication**, following modern security protocols and DevSecOps principles.

## PROJECTS

**Travel Genie (React, Node.js, AI/LLM, Redis, MongoDB, TailwindCSS, Google OAuth, Google Maps API) | [\[link\]](#)**

- Designed and developed an **AI-driven travel planning application** that personalizes itineraries by gathering user preferences such as destination, duration, budget, and group size, resulting in tailored trip recommendations.
- Integrated advanced technologies including **React, Node.js, large language models (LLMs), Redis, MongoDB, and TailwindCSS** to ensure a scalable, responsive, and user-friendly experience.
- Implemented **Google OAuth** for secure authentication and **Google Maps API** for dynamic places autocomplete, enhancing user engagement and streamlining trip planning.

**MediInsight (Python, Django, MySQL, NLP, Machine Learning)**

- Engineered an **automated healthcare analytics system** leveraging **natural language processing** and **machine learning**, achieving **76% illness identification accuracy**.
- Designed secure, extensible **APIs** to interface predictive models with external health systems.
- Developed an automated healthcare analytics system using **NLP** and **ML**, achieving **76% illness classification accuracy** Applied **data anonymization and database security best practices** to support **regulatory compliance** and patient privacy.

## Proximate (HTML, CSS, JavaScript, Firebase, Location Services, PayPal Payment Processor) | [\[link\]](#)

- Engineered a **real-time proximity-based chat application**, enabling user discovery and interaction within dynamic geographic bounds.
- Implemented **end-to-end encryption** and OAuth-based login for robust security and user trust.
- Added **location tracking** and **payment processing with PayPal**, supporting secure in-app transactions.

## SKILLS

---

**Languages:** Python, Java, C++, JavaScript, TypeScript, SQL

**Frameworks & Tools:** FastAPI, Flask, Node.js, Express.js, React.js, JUnit, Tailwind CSS

**System Design & Architecture:** Microservices, Scalable API Design, Authentication (OAuth2, JWT)

**Databases:** MySQL, MongoDB, Redis, Firestore

**Cloud & DevOps:** AWS, GCP, Docker, Kubernetes, Git, GitHub, GitLab CI/CD

**Concepts:** Algorithms & Data Structures, System Design, Concurrency, Message Queues, Kafka, Data Analytics, RESTful APIs, GraphQL

## RESEARCH & PUBLICATIONS

---

### Proposed Model of Speech Recognition using MFCC and DNN | [\[view paper\]](#)

**International Journal of Engineering Research & Technology (IJERT) · May 10, 2020**

- **Project Objective:** The project focuses on converting human speech into text using speech recognition technology to enable voice-based input for applications and facilitate efficient human-machine interaction.
- **System Workflow:** The system architecture involves fetching speech data, preprocessing it using the Polygon smoothing algorithm, extracting features with MFCC, and performing classification using models like SVM and DNN.
- **Comparative Analysis:** While SVM is used for classification, the study shows that DNNs offer greater precision and accuracy, especially with large datasets, leading to a more robust speech recognition system.

### An E-Health Patient Management System | [\[view paper\]](#)

**Grenze International Journal of Engineering and Technology (GIJET) · Jul 14, 2021**

- **Remote Medical Access:** The system enables patients to connect with doctors virtually, share symptoms via voice input, and receive personalized healthcare insights through an intuitive web application.
- **Symptom Analysis Using NLP:** Patient symptoms are recorded, stored in a database, and analyzed using natural language processing (NLP) to ensure smooth and accurate communication with healthcare professionals.
- **AI-Powered Diagnosis Support:** A machine learning model predicts potential diseases based on symptoms, streamlining the diagnosis and treatment process and improving the overall accessibility and efficiency of healthcare delivery.