

# PREETAM MASKE

San Jose, CA | (669)-340-6332 | [preetam.a.maske@gmail.com](mailto:preetam.a.maske@gmail.com) | [linkedin/preetam-maske/](https://www.linkedin.com/in/preetam-maske/) | [github.com/Preetam3620](https://github.com/Preetam3620)

## EDUCATION

### Master of Science, Computer Software Engineering

Aug 2024 - May 2026

San Jose State University, San Jose, California

Coursework: Systems Software Engineering, Cloud Technologies, Distributed Systems, Scalable System Design

### Bachelor of Engineering, Computer Engineering

Jul 2018 - Jul 2022

Savitribai Phule Pune University, India

Coursework: Data Structures & Algorithms, Object Oriented Programming, Database Management, Cloud Computing

## TECHNICAL SKILLS

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

**Backend & Frameworks:** RESTful Web Services, Microservices, Node.js, Flask, FastAPI

**Databases & Messaging:** MySQL, MongoDB, Redis, Supabase, Firebase, Apache Kafka

**Cloud & DevOps:** AWS, Docker, Terraform, CI/CD, Bamboo, CircleCI, Git, Shell Scripting, Linux, Windows

**AI & Data:** LangGraph, LangChain, RAG Pipelines, Agentic AI, Pinecone, ChromaDB, Deepgram

## WORK EXPERIENCE

### Software Engineer, Volkswagen Group Technology Solutions, Pune, India

Aug 2022 - Jul 2024

- Owned the Spare Parts Portal end-to-end; from requirements through production, serving 10,000+ enterprise users, reducing stakeholder back-and-forth by 30% through iterative delivery and documented API contracts.
- Eliminated SAP posting bottlenecks by introducing Kafka to decouple OCR, validation, posting stages; designed topics, partitions, and retention policies that cut processing pipeline latency by 40% under peak traffic spikes.
- Engineered microservices backed by MySQL with Redis caching, added targeted indexes and query optimizations that improved P95 API response time by 35% under real production load.
- Integrated Computer Vision OCR into invoice processing workflows using Flask and MongoDB, automating structured data extraction and reducing manual data entry effort by 60% for the finance operations team.
- Automated CI/CD pipelines using Atlassian Bamboo and shell scripting, reducing average release cycle time by 50% and enabling repeatable, low-risk deployments across both AWS EC2 and on-prem environments.
- Configured AWS security groups, IAM policies, and EC2 health checks for production Flask services, hardening the deployment environment and maintaining zero security incidents across a 2-year production run.

### Software Developer Intern, Simba Developers

Jan 2021 - Jul 2021

- Delivered a real-time sensor monitoring dashboard using Node.js and REST APIs, streaming live device data with sub-second update latency and reducing manual reporting effort for operations teams by 70%.
- Shipped core features for a cross-platform application in Flutter and Dart, improving input-to-render latency by 25% through UI profiling and targeted rendering optimizations on Android and iOS.
- Standardized API contracts and led code reviews across a 6-person team, catching integration defects pre-merge and reducing post-release bug reports by 40% over the internship period.
- Researched and compared SQL vs NoSQL trade-offs for a home automation backend, producing a documented architecture proposal that the team adopted, cutting estimated future service integration time by 2 weeks.

## PROJECTS

### BookTable: Restaurant Reservations Platform

Jan 2025 - Mar 2025

- Architected role-based RESTful APIs in FastAPI supporting Customer, Manager, and Admin workflows, reducing booking conflict errors by 90% through input validation and idempotent request handling.
- Tuned relational schema indexes and added Redis caching layer, reducing average booking query latency by 35% and sustaining throughput under concurrent reservation load.
- Deployed containerized services on AWS EC2 behind an ALB with Terraform-provisioned auto-scaling, supporting up to 10,000 concurrent users with zero-downtime blue-green releases.

### Rufus: AI-Powered Shopping Assistant

Sept 2025 - Oct 2025

- Implemented a LangGraph multi-agent workflow routing conversational, text, and image queries through GPT-4o, improving recommendation relevance by 40% over single-pass retrieval approaches.
- Built a Pinecone semantic retrieval pipeline with multi-stage re-ranking and business-rule boosting, reducing irrelevant product results by 45% across diverse catalog searches.
- Designed stateful FastAPI backend with Redis session persistence, maintaining context across multi-turn interactions and cutting average response latency to under 800ms at P95.

### AlphaRescue: AI-Driven Emergency Dispatch System

June 2025 - Sept 2025

- Implemented a VAPI voice assistant powered by Gemini that transcribed live distress calls, filtered false alarms, and generated structured incident briefs in under 3 seconds end-to-end.
- Architected a network of specialized Fetch.ai uAgents on AgentVerse that collectively assessed severity, classified response type, and queried a Supabase RAG pipeline, reducing manual dispatch decision time by 70%.
- Launched a real-time React and Mapbox dashboard visualizing active incidents and responder locations, cutting situational awareness lag for control teams from minutes to under 10 seconds.

## ACHIEVEMENTS

UC Berkeley AI Hackathon 2025 - Best AI Agent Deployment, Awarded for AlphaRescue, June 2025

ISTE Hack-SPRINT v3.0 2021, a 36-hour hackathon - Business Impact Award, September 2021