

Professional Summary

AI/ML Software Engineer with an MS in CS and 2+ years of experience specializing in Agentic workflows and scalable AI backends. Proven track record of reducing operational costs by 90% through LLM automation and cloud-native deployment (GCP/FastAPI).

Experience

AllCheer (Fateh and Sudha Inc), San Francisco, CA

Software Developer (Full-time)

Mar 2025 – Present

- Engineered a fully automated Release of Information (ROI) system using a **LangChain AI agent** to orchestrate document generation, e-signature collection, and secure data sharing, saving **\$20K/year** and **reducing processing time by 60%**.
- Deployed a HIPAA compliant production-grade **FastAPI backend** to ingest real-time data from Rethink BH, building out an observability stack on **Google Cloud Run** that **cut costs by 90%** and eliminated 24-48h latency.
- Developed an internal dashboard integrating a GPT-based NLP API for **named-entity recognition (NER)** and text summarization to structure unformatted therapist notes, improving **data accuracy to 92%**.

AllCheer (Fateh and Sudha Inc), San Francisco, CA

Software Developer Intern

Jun 2024 – Dec 2024

- Architected and automated a logistics engine to schedule trips for the staff using **Google Route / Distance Matrix API**.
- Validated ROI automation feasibility through **Make.com** prototyping, leading to the full-scale production implementation currently in use.

Pinnacle Consulting LLC (PCON Utilities Pvt Ltd), Bhubaneswar, India

Junior Software Developer

Jan 2022 – Aug 2023

- Integrated **ArcGIS** mapping with REST APIs and SQL Server for a \$75M utility provider in North Carolina to visualize infrastructure and sync real-time updates from **mobile field apps**, reducing oversight errors by **30%**.
- Delivered 3 **ERP systems** using **Angular** and **Node.js** for internal team of 100+ staff to manage **timesheets, communication, and KPI reporting**; implemented **JWT authentication**, automated invoicing, improving operational efficiency by **25%**.

Technical Skills

- Languages:** Python, JavaScript, TypeScript, SQL
- AI & ML:** LangChain, LangGraph, RAG, PyTorch, TensorFlow, Prompt Engineering, NLP, HuggingFace
- Backend & Cloud:** FastAPI, RESTful API, Node.js (Express), Docker, Kubernetes, Google Cloud, AWS (S3, Lambda)
- Databases & Tools:** PostgreSQL, MongoDB, Redis, Pinecone (Vector DB), Git

Projects

bedtime.ai – Multi-Modal AI Storytelling Platform

A multi-modal AI pipeline that converts drawings to narrated stories; utilized **EfficientNet** for image feature extraction, a **fine-tuned phi-3 LLM** for story generation, and **coqui-ai TTS** for voice cloning.

Agentic Job Helper:

Stateful multi-agent system using **LangGraph** and **GPT-4o** for automated job search and application tracking; managed states with **FastAPI** and **Supabase**.

RAG for Document Intelligence

Built a RAG pipeline for document Q&A; implemented semantic chunking and embedding generation to index PDFs in **Pinecone**, enabling a **LLM** interface with 35% higher relevance over baseline keyword search.

Education

University of California, Riverside

Master of Science in Computer Science

GPA: 3.61

2023 - 2025