

SUKRUTH MOTHAKAPALLY

Dublin, CA | +1 603.688.4686

Sukruth0402@gmail.com | linkedin.com/in/sukruthkm | github.com/Sukruthmothakapally | sukruthmothakapally.me

Work Experience

AI Software Engineer

Feb 2025 - Present

[Predicta.bio](#), Boston, MA (Python, LLMs, ML, FastAPI, MongoDB, Azure, Airflow, Playwright, Cursor)

- **Founding engineer** leading **end-to-end** development of scalable **multi-agent AI microservices** for **Biotech's MVP**.
- Contributed to **ML pipelines** integrating **multi-modal** data, graph and **vector DB** to enhance protein expression prediction.

AI Software Engineer

Jan 2024 - Aug 2024

[PipeIQ.ai](#), Palo Alto, CA (Python, LLMs, RAG, FastAPI, AWS, Postgres, Node.js, OpenAI)

- As part of founding team, collaborated directly with **CEO** in a **0→1** initiative to design and develop **4** core **microservices** for a **SaaS MVP** using FastAPI/Node.js/AWS Lambda/Kafka/Docker, accelerating time-to-market by a **quarter**.
- Contributed to building a **multi-agent AI system** (OpenAI/fine-tuning/LangChain/Selenium) for **automating** and **hyper personalizing** email generation, lead, intent, and campaign creation, boosting **sales** conversions and acquiring **10+** customers.
- Developed an **AI chatbot** (RAG/Cache/FAISS/WebSocket) serving as a Customer **Onboarding Assistant**, delivering real-time sales insights and slashing onboarding time from **hours to minutes**.
- Designed **8** database schemas for sales-marketing workflows and led **agile** migration from monolith to microservices backend, integrating **AWS** services with **CI/CD** pipelines via GitHub Actions/Terraform, streamlining weekly **release cycles**.

Graduate Teaching Assistant

Jan 2023 - Dec 2024

[Northeastern University](#), Boston, MA (Python, ML, Object Oriented Programming)

- **Co-created** and graded assignments, mentoring **40+** students in **Data Structures**, **Algorithms** and **Data Science** courses.

Systems Engineer

Dec 2020 - June 2022

[Infosys \(Intel\)](#), Bangalore, India (Python, C++, MySQL, CI/CD, Automation, Testing, TDD)

- Automated Intel **GPU/SSD** driver deployments via C++ CLI tools and **Ansible** on Linux, configuring **10K+** enterprise systems.
- Built a Python **automation** library using **TDD**, writing **50+** **SSD** firmware tests and improving test throughput by **15%**.
- **Re-engineered** driver updates using Python/Shell **parallel scripting**, slashing deployment time by **30%** for **1M+** users.
- Optimized **CI/CD** pipelines (Jenkins/Git) with automated unit/integration tests, reducing driver build failures by **7%**.

Full-Stack Developer Intern

Feb 2020 - Mar 2020

[Infosys](#), Mysore, India (JavaScript, MERN, Agile)

- Developed **4** REST APIs (Express.js/MongoDB) using aggregation pipelines and **indexing**, enhancing latency by **10%**.
- Built **real-time** analytics dashboard (React/D3.js) with code-splitting and **DOM** optimization, reducing load time by **9%**.

Software Developer Intern

Jul 2018 - Aug 2018

[Bosch](#), Bangalore, India (Python, Node.js, REST, gRPC, AWS, Docker)

- Enhanced microservices with **JWT** authentication and **session** management, improving API security for shipment tracking.
- Implemented **Redis caching** and optimized SQL queries using indexing, reducing data retrieval time by **30%**.

Personal Projects

[WhatsApp Multi-Agent and Multi-Modal AI Assistant](#) (LangGraph, LLMs, RAG, STT/TTS, ITT/TTI, FastAPI, Prefect, Cron)

- Developed a scalable multi-agent system with graph-based routing, integrating WhatsApp and Google Workspace APIs.
- Built asynchronous workflows and specialized extraction pipelines for real-time processing of text, audio, and images.

[Personalized AI Job Tracker – Full-Stack App](#) (Python, LLM, FastAPI, JS, Postgres, AWS, Docker, Prefect)

- Developed a **Chrome extension** with **multi-agent AI** microservices to automate job tracking, deployed on AWS with Docker.

[Custom LLM & GPT Fine-Tuning with Multi-Agent AI](#) (PyTorch, LLM, LoRA, RLHF, NLP, Deep Learning)

- Engineered a GPT based LLM with **AI agents** using **LoRA**, **RLHF**, and instruction backtranslation for aligned responses.

[Audio Emotion Classifier with Deep Learning and Parallel Computing](#) (Python, PyTorch, CNN, NLP)

- Built an audio emotion classifier using **DL**, **NLP**, and **parallel** computing to efficiently detect complex emotion patterns.

[AI-based Stack Overflow Full-Stack App](#) (Python, LLM, Airflow, Docker, FastAPI, CI/CD, GCP)

- **Led** a team of **3** to build a **containerized full-stack app** integrating **AI agents** into **Stack Overflow**, deployed on GCP.

Technical Skills

- **Language** | **Databases**: Python, Java, JavaScript, SQL | PostgreSQL, MySQL, MongoDB, Redis, Pinecone, Elasticsearch.
- **AI** | **LLM**: Multi Agent System, RAG, MCP, Multi-Modal, tool call, prompting, fine-tune, LangChain, LangGraph, Cursor.
- **ML** | **NLP**: TensorFlow, Sklearn, PyTorch, CNN, GAN, Sagemaker | BERT, Transformers, Hugging Face, spaCy.
- **Backend**: FastAPI, Django, Node.js, RESTful APIs, GraphQL, gRPC, Microservices, Kafka, System Design, SDLC.
- **Cloud** | **CI/CD**: AWS, GCP, Azure | Git, GitHub Actions, Jenkins, Terraform, Docker

Education

[Northeastern University](#), Boston, MA - **Master of Science** in Computer Software Engineering. (3.8 / 4.0)

Dec 2024

Relevant Courses: Data Structures and Algorithms, Data Science, Advances in LLMs, Cloud, Big Data, Parallel Computing.

[Visvesvaraya Technological University](#), Bangalore, India - **Bachelor of Engineering** in Computer Science and Engineering