SUKRUTH MOTHAKAPALLY

United States | +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, n8n)
- Founding engineer **developing** and **deploying** high throughput microservices, RESTful APIs, and **distributed systems**, integrating a **multi-agent AI** backend for a critical **biotech MVP**, ensuring robust system reliability.
- Designed and implemented **scalable data ingestion** pipelines and backend orchestration using **AI/ML** models, feature engineering, and **automated** workflows to enable efficient **protein data** processing.

AI Software Engineer   Jan 2024 - Aug 2024
PipeIQ.ai, Palo Alto, CA (Python, LLMs, RAG, FastAPI, AWS, Postgres, Node.js, OpenAI)
- Founding engineer of a **fast-paced** startup, collaborated closely with **CEO** in a **0→1** initiative to develop core **SaaS** backend services, leveraging FastAPI, AWS Lambda, Kafka, and **container** orchestration to drive scalable **cloud-native** solutions.
- Engineered **multi-agent AI** workflows leveraging **RAG** pipelines, fine-tuned **LLMs**, LangGraph, and Selenium to automate and **hyper-personalize** email generation, lead processing, and campaign orchestration, enabling a natural language interface (**NLI**).
- Developed an **AI chatbot** with **Tool Calling** (RAG/Cache/FAISS/WebSocket) serving as an **automated** Customer **Onboarding Assistant**, delivering **real-time** sales insights and slashing onboarding time from **hours to minutes**.
- **Led** design and optimization of relational **data models** and transitioned legacy monolithic backend to **microservices** architecture, integrating automated **CI/CD** pipelines (GitHub Actions/Terraform) for continuous deployment and **monitoring**.

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 secure** RESTful 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, ensuring **data integrity** and user privacy.
- 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 (DAG, LangGraph, LLM, RAG, STT/TTS, ITT/TTI, FastAPI, Prefect, Cron)
- **Graph**-based routing, webhook, real-time text/audio/image processing, **WhatsApp and Google APIs**, OAuth, Docker.

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, Hugging Face)
- **Fine-tuned** 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, LangGraph, Cursor, n8n, vibe code
- **ML | NLP:** TensorFlow, Sklearn, PyTorch, CNN, GAN, Sagemaker | BERT, Transformers, Hugging Face, spaCy, NLTK
- **Backend:** FastAPI, Django, Node.js, Express.js, GraphQL, gRPC, Microservices, Kafka, System Design, SDLC, Agile
- **Cloud | CI/CD**: AWS, GCP, Azure | Git, GitHub Actions, GitLab, 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