Sampath Garimella

628-309-5036 | sampath.garimella.dev@gmail.com | linkedin.com/in/sampathgarimella | github.com/SampathGarimella

SUMMARY

Full stack and AI Developer with 4+ years of experience building scalable cloud apps and AI frameworks. Skilled in Python, Swift, React, and RAG pipelines. Built internal automation tools adopted across engineering teams, delivered under tight schedules, and scaled to thousands of users.

EDUCATION

University at Buffalo – *MS, Engineering Science,* Buffalo, NY   Jan 2022 – May 2023
Graduate Teaching Assistant, *Machine Learning with Search Engines*
Courses: Machine Learning, Deep Learning, Advanced NLP, Search Engines, ML for Large Datasets, Multimodal ML, Computer Vision

KL University - *BTech, Computer Science,* India   May 2017 – Apr 2021

TECHNICAL SKILLS

Programming Languages:	Python, JavaScript, TypeScript, Swift, Java, Go, C++, C#, .NET Core
Frontend:	React, Next.js, Vue.js, Angular, React Native, HTML5, CSS3, Tailwind CSS, Bootstrap
Backend & Architecture:	Node.js, Django, Flask, Spring Boot, GraphQL, REST, gRPC, WebSocket, Microservices, Event-Driven Architecture
AI/ML:	LLMs, RAG, Transformers, NLP, Computer Vision, Hugging Face, LangChain, PyTorch, TensorFlow, scikit-learn, Pandas, NumPy
Data & Search:	Elasticsearch, Pinecone, FAISS, Chroma, PostgreSQL, MySQL, MongoDB, DynamoDB, Redis, Cosmos DB, Kafka
Cloud & MLOps:	AWS (SageMaker, Bedrock, EC2, RDS, S3, Lambda), Azure (AI Studio, Service Bus, Event Hub), GCP (BigQuery), Docker, Kubernetes, Terraform, MLflow, Jenkins, GitHub Actions
Security:	RBAC, OAuth 2.0, JWT, IAM hygiene, TLS, audit logging, OWASP Top 10 awareness
Observability & QA:	Datadog, Grafana, New Relic, Splunk, SonarQube, Postman, Cypress, Selenium, JUnit, Jest
AI-Assisted Development:	Copilot, ChatGPT, Claude Code, Cursor for testing, refactoring, and documentation support

EXPERIENCE

Mastery Logistics Systems - Atlanta, GA   Oct 2024 – Present
Full Stack Developer (Generative AI)

- Design and deploy ML-driven route optimization solutions, increasing operational efficiency by 25% and saving $100K annually
- Build and maintain LLM-powered logistics reporting tools with LangChain, Pinecone, and FAISS, enabling context-aware search and summarization
- Develop and optimize RAG-based recommendation modules for shipment planning, reducing decision time by 60%
- Architect and secure GraphQL/REST APIs serving 500K+ daily requests with 99.99% uptime, supporting scalable operations
- Deploy and monitor AI workflows on AWS Bedrock and SageMaker, optimizing performance and cost efficiency
- Delivered compliance-critical tools on tight deadlines, adopted by multiple logistics teams, driving measurable impact.
- Developed internal AI tooling with Python, React, and Swift to streamline logistics workflows and improve user adoption
- Used code assistants (Copilot, ChatGPT/Claude Code, Cursor) to speed up implementation and test coverage while keeping code reviews strict
- Containerized services with Docker/Kubernetes, built CI/CD pipelines in GitHub Actions, and monitored with Datadog, Grafana, and Splunk.

UPMC - Pittsburgh, PA   Dec 2023 – Sep 2024
Full Stack Developer

- Built HIPAA-compliant data platforms with Django REST Framework for 90K+ patients
- Developed React/Next.js reporting modules, improving provider satisfaction scores by 30%
- Collaborated with clinicians and UX teams to design intuitive healthcare dashboards.
- Integrated NLP pipelines to analyze clinical text, reducing manual review time by 50%
- Worked across engineering and clinical teams to ensure successful adoption of tools by thousands of healthcare providers
- Added role-based access, audit trails, and TLS/IAM-secure data handling patterns to protect patient data

Cigna - Overland Park, KS   Jan 2022 – Nov 2023
Full Stack Developer

- Developed Kafka-based analytics pipelines and integrated DynamoDB for sub-50 ms query latency
- Developed healthcare features with React Native + Spring Boot, increasing adoption by 15%
- Authored unit/integration tests and targeted refactors, reducing bugs found after release
- Strengthened security with OAuth 2.0 and JWT, and IAM hygiene practices ensuring HIPAA compliance

Tech Mahindra - India   Jan 2021 – Dec 2021
Web Developer

- Rebuilt a legacy front end into a reusable Vue.js component library, modernizing the UI and improving load speed by 20%
- Implemented Selenium/Jenkins pipelines, reducing post-release defects by 25%

PROJECTS

AI-Powered Logistics Optimization *(Generative AI, LLM, RAG, Vector Search, AWS Bedrock, SageMaker)*   Oct 2024 – Present

- Built a generative AI system for shipment scheduling and routing with vector search, cutting planning time by 60% and boosting throughput by 25%
- Integrate RAG pipelines with logistics microservices for real-time decision-making and adaptive routing
- Ship with Docker/Kubernetes, CI/CD checks, and useful service dashboards to support fast iteration

Document Intelligence & Summarization Tool *(LLM, RAG, Transformers, NLP*)   Oct 2024 – Present

- Develop an NLP platform to summarize shipment contracts and compliance documents using transformer-based NLP models and RAG pipelines
- Reduce manual document review time by 55% and improve compliance turnaround speed
- Used assistants to bootstrap tests and docs, handled sensitive content with care in storage and logs
- Delivered automation framework that accelerated compliance workflows and was adopted org-wide

Intelligent Visual Search Engine *(Multimodal Search, ResNet, Elasticsearch, NLP)*   Jan 2022 – May 2023

- Designed a multimodal search engine enabling image-to-image and text retrieval using ResNet embeddings and cosine similarity in Elasticsearch
- Combined visual embeddings with NLP-based caption search for hybrid retrieval, improving result relevance by 35% over keyword-only search