RAKSHEKA RAJAKUMAR

raksheka.me/ | linkedin.com/in/raksheka/ | github.com/rakshekaraj | rakshekaraj@gmail.com | +1 (323) 646-3734 | Los Angeles, CA

EDUCATION

University of Southern California — **Los Angeles, CA**
Master of Science in Electrical & Computer Engineering- Machine Learning & Data Science — August 2023-May 2025
- **Courses**: Data structures and algorithms, Information Retrieval, Cloud Computing, Deep learning systems, Probability

Anna University - Coimbatore Institute of Technology — **Coimbatore, India**
Bachelor of Engineering in Electronics and Communication – CGPA: 9.1/10 — July 2019-May 2023
- **Courses**: Data Networks, Digital Image Processing, Machine learning, Statistics

TECHNICAL SKILLS

Languages: Python, C++, C#, SQL, HTML, CSS, JavaScript
Web Technologies: React.js, Node.js, Angular, FastAPI, Flask, Streamlit, MERN Stack, LAMP stack, Web APIs
Database & Vector storage: MongoDB, MySQL, PostgreSQL, DynamoDB, Pinecone
Cloud platforms: Docker, Kubernetes, Kubeflow, Linux, AWS
AWS Services: SageMaker, S3, EC2, Bedrock, CloudWatch, API Gateway, Lambda, DynamoDB, IAM
DevOps, Automation & Monitoring: Git, Jenkins, JIRA, Agile (Scrum), Splunk, Bash, PowerShell
Frameworks & MLOps: TensorFlow, Keras, PyTorch, Pyspark, Scikit Learn, NumPy, Matplotlib, Langchain, LlamaIndex, ML Flow, Hadoop

EXPERIENCE

WorkUp — **Los Angeles, CA**
Machine Learning Engineer — May 2024-July 2024
- Constructed a personalized state-of-the-art job recommendation engine for WorkUp to match job seekers with relevant job postings, enhancing match quality and user engagement through NVIDIA Merlin's Two-Tower architecture on AWS SageMaker
- Translated product requirements into a scalable ML pipeline by mapping user profiles to job embeddings, enabling real-time recommendations
- Engineered feature transformations with NVTabular and evaluated impact using ranking metrics (Recall@K, NDCG), for 15% higher precision
- Leveraged MongoDB to retrieve user data and application history at runtime, compute user embeddings via query tower, and return Top-K jobs

Kanini Software Solutions — **Chennai, India**
Software Engineering Intern — February 2023-June 2023
- Led the design and development of a modular Loan Product Management dashboard that streamlined banking workflows across deposits, loans, credit cards, and life insurance, featuring scalable CRUD architecture, reducing operational processing time by 40%
- Created normalized schemas, ETL pipelines, and dynamic UI modules; deployed APIs on AWS EC2 with Docker and Kubernetes
- Optimized data handling and transaction workflows by building modular micro services and event-driven architecture, achieving 70% boost in platform scalability and user engagement
- Collaborated cross-functionally with QA, product managers, and senior engineers to deliver features under Agile sprints; contributed to internal documentation and demo walkthroughs for client onboarding

MSAII — **Chennai, India**
Software Engineering Intern — November 2022-January 2023
- Designed micro services infused backend (FastAPI, Celery, RabbitMQ) to handle bill uploads and queue OCR/ML tasks, with 99% uptime
- Fine-tuned LayoutLMv3 model (PyTorch, Hugging Face), achieving 86% field-level accuracy, which reduced manual extraction effort by 60%
- Built a React dashboard (React Query, WebSockets), capturing user corrections on 85% of extracted fields to feed into continuous retraining
- Containerized all services with Docker, deployed on Kubernetes (AWS ECS Fargate), and implemented CI/CD pipelines (GitHub Actions) that cut deployment time by 30%

MITACS — **Waterloo, Canada**
Globalink Research Intern — June 2022-September 2022
- Assembled and modeled k8 rovers for a project sponsored by Canadian Space Agency, demonstrated usage through seminars for PhD candidates, under Dr Julie Mueller and Inksmith Technologies
- Researched and applied RCNN models to increase speed in object detection in rover prototypes by 15%

PROJECTS

AirPulse Delhi: Distributed Cloud-Native AQI Intelligence Platform
- Devised serverless real-time AQI analytics system on AWS Lambda, DynamoDB, API Gateway, SNS; reduced latency by 25% through event-driven processing and scalable NoSQL storage
- Orchestrated modular micro services for ingestion pipelines, ML-driven AQI trend forecasting, personal exposure profiling, dynamic alert rule generation, for improved active-user time by 10%
- Implemented React.js frontend via AWS Amplify, Route 53, visualizing multi-zone AQI heat maps, historical spatiotemporal trends and user maps
- Embedded lightweight ML inference inside Lambda functions for spatiotemporal AQI analytics and real-time API inferencing

EDQty- Multimodal learning Web Application
- Spearheaded full-stack web application leveraging Flask (backend), React (frontend), and MongoDB (NoSQL database) to deliver low-latency, multimodal user interaction workflows
- Constructed vector search pipeline with Pinecone DB, enhancing embedding retrieval operations & improving query performance by 25%
- Integrated Llama 3.2 LLM to enable 20% faster Retrieval-Augmented Generation (RAG) pipelines across retrieval, multilingual translation, summarization, and multimodal captioning tasks

Personalized Federated Learning: Contribution-Based Aggregation for 3D Brain Tumor Segmentation
- Built a privacy-preserving federated learning pipeline for distributed image segmentation
- Engineered transformer encoded UNet architecture having a global encoder and localized decoders to outperform centralized model by 6%
- Introduced Contribution Factor as a weighting mechanism allowing more diverse and high-performing clients to have greater influence on global model during weight averaging, and validated approach on BraTS 3D dataset

PUBLICATIONS

- Performance Analysis of CNN Architectures in Multi-label Image classification, Proceedings **IJCA** 184(48):14-18, February 2023
- Assessment of ML Algorithms for Predicting Campus Placements, Proceedings of **ICMCSI** 2023 (pp. 221-231 Springer)