Sarabesh Neelamegham Ravindranath

sarabeshnr@gmail.com | +1-930-333-4115 | linkedin.com/in/sarabesh | github.com/sarabesh

Professional Summary

Platform-focused ML Infrastructure / MLOps Engineer with 4+ years of experience across DevOps, cloud infrastructure, Kubernetes platforms, and production ML workflows. Builds reliable CI/CD systems, dataset and model-versioning pipelines, training and batch-inference workflows, release automation, and observability tooling across AWS, Kubernetes, Terraform, Docker, GitLab CI/CD, DVC, and MLflow. Proven track record improving deployment velocity, reproducibility, platform reliability, and operational efficiency for cloud-native and ML systems.

Work Authorization: Currently on F-1 STEM OPT with EAD valid through June 2028. H-1B sponsorship will be required thereafter for continued work authorization or change of status.

Technical Skills

Programming and Frameworks: Python, Go, C++, Java, Groovy, Flask, FastAPI, Spring Boot
ML/MLOps Platforms: DVC, MLflow, SageMaker, Prefect, Airflow, model registry, experiment tracking, dataset versioning, batch inference
ML/AI Systems: Computer Vision, RAG, Vector Search, NLP, Reinforcement Learning, model deployment, model monitoring, drift detection
ML Libraries: PyTorch, TensorFlow, Scikit-learn, Hugging Face, LangChain, Open3D, TensorRT
Cloud and Infrastructure: AWS Lambda, ECS, EKS, RDS, S3, Step Functions, IAM, VPC, Kubernetes, Docker, VMware Tanzu, NFS, SLURM
DevOps, CI/CD, and IaC: Terraform, CloudFormation, Helm, Jenkins, Concourse, GitHub Actions, GitLab CI/CD, Argo CD
Databases and Storage: MySQL, DynamoDB, Neo4j, Qdrant, LMDB, PuppyDB
Certifications: AWS Certified Solutions Architect - Associate (SAA), Certified Kubernetes Administrator (CKA), Certified Kubernetes Security Specialist (CKS), HashiCorp Terraform Associate, NVIDIA-Certified Associate: AI Infrastructure and Operations

Experience

Research Assistant - Indiana University School of Optometry, Bloomington, IN — Jan 2025-Present

- Collaborate with faculty on the Walking Project, building reproducible 3D terrain-reconstruction and complexity-modeling workflows from data collected across 10+ trails and 8 subjects using RealityCapture.
- Reconstructed 10+ scenes from binocular and monocular imagery, creating terrain-analysis assets for scene-understanding and surface-complexity experiments.
- Plan field data collection and evaluate multimodal capture tooling including Meta Aria eye-tracking glasses, drone/overhead cameras, GPS trackers, and Shadow IMU body suits.
- Train and evaluate models on the Nymeria dataset for single-image camera-orientation detection using IU supercomputing workflows, SLURM job orchestration, environment modules, and distributed training as needed.

Emerging Tech Intern, AI Specialist - Musco Lighting, Des Moines, IA — May 2024-Aug 2024

- Developed a modular MLOps pipeline with DVC and GitLab CI for dataset versioning, validation, retraining, and model promotion, improving iteration speed by 40%.
- Implemented YOLOv5 training and batch-inference workflows integrated with MLflow for experiment tracking, artifact management, model comparison, and performance visibility.
- Designed event-driven dataset synchronization between NFS and S3, reducing data latency by 80% and improving availability for downstream ML workflows.
- Enhanced production ML observability with MLflow dashboards, drift signals, and data change alerts to strengthen model lifecycle governance.

DevOps Engineer, MTS II - VMware, Bangalore, India — Feb 2022-Jun 2023

- Contributed to the VMware Tanzu Application Platform (TAP) release team, automating multi-cloud installation and validation pipelines across EKS, AKS, and TKGS.
- Designed and implemented Go-based E2E validation suites, increasing automated test coverage from 20% to 75% across platform-release workflows.
- Integrated CI/CD workflows into Helm-based Kubernetes infrastructure, improving release reliability and reducing manual validation overhead.
- Partnered with platform, QA, and release teams to optimize Kubernetes deployments, version control integration, automated environment validation, and release readiness.

Cloud Engineer - Presidio, Chennai, India — May 2019-Jan 2022

- Architected and deployed enterprise CI/CD pipelines using Jenkins and GitHub Actions, implementing Infrastructure as Code via Terraform and CloudFormation.
- Built scalable AWS infrastructure for clients including ViacomCBS, StatsPerform, and WEX, supporting cloud-native deployment, cost optimization, and operational reliability.
- Led migration of Jenkins workers to Kubernetes pods, improving resource efficiency and reducing build times by 35%.
- Developed internal DevOps libraries integrating Slack and Sumo Logic APIs for real-time alerting and monitoring, reducing mean time to recovery (MTTR).

Projects

PuppyDB - Minimal Vector Database *May 2025-Present*
Developed a lightweight vector database using LMDB for persistent storage and custom HNSW indexing for approximate nearest-neighbor retrieval. Implemented metadata management, brute-force and graph-search modes, and persistent storage for large-scale embeddings. GitHub

Sentiment Analysis - Fine-Tuning, Deployment and Monitoring *Jan 2024-Apr 2024*
Built an end-to-end MLOps workflow for model deployment and lifecycle management. Fine-tuned BERT on the Sentiment140 dataset, served the model with FastAPI, and deployed on Kubernetes (Minikube) using Docker. Implemented batch prediction jobs, data drift detection, Slack-based alerting, and automated retraining workflows using CronJobs. GitHub

HybridRAG - Graph vs Vector RAG Benchmark *Aug 2024-Dec 2024*
Benchmarked GraphRAG (Neo4j) against Vector RAG (LangChain + Ollama) architectures for LLM retrieval systems. Evaluated retrieval latency, contextual precision, scalability, and infrastructure tradeoffs across hybrid knowledge pipelines. GitHub

GridWorld-PPO - Path and Route Planning with RL *Jan 2025-May 2025*
Implemented reinforcement-learning agents using PPO in a 2D grid-world simulator with dynamic obstacles. Integrated reward shaping and imitation learning to improve convergence, visualized real-time trajectories, and benchmarked policy stability across randomized maps. GitHub

Neural-Recon - Image Reconstruction from fMRI *Sep 2024-Dec 2024*
Reconstructed natural scenes from brain activity using latent diffusion models and CLIP embeddings. Mapped fMRI signals to latent representations aligned with semantic image features and replicated experiments using the Natural Scenes Dataset (NSD). GitHub

Mono-Cam DeepVO - Monocular Visual Odometry *May 2024-Jul 2024*
Reimplemented DeepVO for monocular visual odometry, combining CNN (FlowNet-based) feature extraction with LSTM temporal modeling. Trained on the KITTI odometry dataset to predict 6-DoF camera motion, achieving stable trajectory estimation and reduced drift error. GitHub

Education

Master of Science in Data Science
Indiana University Bloomington - Luddy School of Informatics, Computing, and Engineering
CGPA: 3.90/4.00
Key Coursework: Applied Machine Learning, Elements of AI, Exploratory Data Analysis, Computer Vision, Applying ML Techniques in NLP, Deep Learning Systems, Autonomous Robotics

Bachelor of Engineering in Computer Science and Engineering
Velammal Engineering College, Anna University - Chennai, India
CGPA: 8.39/10.00